COGNITIVE NEUROSCIENCE

A brain network supporting social influences in human decision-making

Lei Zhang[1,2]* and Jan Gläscher[1]*†

Humans learn from their own trial-and-error experience and observing others. However, it remains unknown how brain circuits compute expected values when direct learning and social learning coexist in uncertain environments. Using a multiplayer reward learning paradigm with 185 participants (39 being scanned) in real time, we observed that individuals succumbed to the group when confronted with dissenting information but observing confirming information increased their confidence. Leveraging computational modeling and functional magnetic resonance imaging, we tracked direct valuation through experience and vicarious valuation through observation and their dissociable, but interacting neural representations in the ventromedial prefrontal cortex and the anterior cingulate cortex, respectively. Their functional coupling with the right temporoparietal junction representing instantaneous social information instantiated a hitherto uncharacterized social prediction error, rather than a reward prediction error, in the putamen. These findings suggest that an integrated network involving the brain's reward hub and social hub supports social influence in human decision-making.

INTRODUCTION

Human decision-making is affected by direct experiential learning and social observational learning. This concerns both big and small decisions alike: In addition to our own experience and expectation, we care about what our family and friends think of which major we choose in college, and we also monitor other peoples' choices at the lunch counter to obtain some guidance for our own menu selection—a phenomenon known as social influence. Classic behavioral studies have established a systematic experimental paradigm of assessing social influence (*1*), and neuroimaging studies have recently attempted to unravel their neurobiological underpinnings (*2*, *3*). However, social influence and subsequent social learning (*4*) have rarely been investigated in conjunction with direct learning.

Direct learning has been characterized in detail with reinforcement learning (RL) (*5*) that describes action selection as a function of valuation, which is updated through a reward prediction error (RPE) as a teaching signal (*5*, *6*). While social learning has been modeled by similar mechanisms insofar as it simulates vicarious valuation processes of observed others (*7*, *8*), most studies only involved one observed individual, and paradigms and corresponding computational models have not adequately addressed the aggregation of multiple social partners.

Despite the computational distinction between direct learning (with experiential reward) and social learning (with vicarious reward), neuroimaging studies remain equivocal about the involved brain networks: Are the neural circuits recruited for social learning similar to those for direct learning? In direct learning, a plethora of human functional magnetic resonance imaging (fMRI) studies have implicated a network involving the ventromedial prefrontal cortex (vmPFC) that represents individuals' own valuation (*9*) and the ventral striatum (VS)/nucleus accumbens (NAcc) that encodes the RPE (*6*). These findings mirror neurophysiological recordings in nonhuman primates showing the involvement of the orbitofrontal cortex and the striatum in direct reward experience (*10*). Turning to social learning, evidence from human neuroimaging studies have suggested similar neuronal patterns of experience-derived and observation-derived valuation, showing that the vmPFC processes value irrespective of being delivered to oneself or others (*7*, *11*). However, recent studies in both human (*12*, *13*) and nonhuman primates (*14*) have suggested cortical contributions from the anterior cingulate cortex (ACC) that specifically tracks rewards allocated to others. Although these findings suggest that direct learning and social learning are, in part, instantiated in dissociable brain networks, only very few studies have investigated how these brain networks interact when direct learning and social learning coexist in an uncertain environment (*15*), and none of them involved groups larger than two individuals.

Here, we investigate the interaction of direct learning and social learning at behavioral, computational, and neural levels. We hypothesize that individuals' direct valuation is computed via RL and has its neural underpinnings in the interplay between the vmPFC and the NAcc, whereas individuals' vicarious valuation is updated by observing their social partners' performance and is encoded in the ACC. In addition, we hypothesize that instantaneous socially based information has its basis in the right temporoparietal junction (rTPJ) that encodes others' intentions necessary for choices in social contexts (*12*, *16*, *17*). To test these hypotheses, we designed a multistage group decision-making task in which instantaneous social influence was directly measured as a response to the revelation of the group's decision in real time. By further providing reward outcomes to all individuals, we enabled participants to learn directly from their own experience and vicariously from observing others. Our computational model separately updates direct and vicarious learning, but they jointly predict individuals' decisions. Using model-based fMRI analyses, we investigate crucial decision variables derived from the model, and through connectivity analyses, we demonstrate how different brain regions involved in direct and social learning interact and integrate social information into individuals' valuation and action selection. In addition, confidence was measured both before and after receiving social information, as confidence may modulate individuals' choices during decision-making (*3*, *18*).

[1]Institute of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. [2]Neuropsychopharmacology and Biopsychology Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, 1010 Vienna, Austria.
*Corresponding author. Email: lei.zhang@univie.ac.at (L.Z.), Twitter: @lei_zhang_lz; glaescher@uke.de (J.G.)
†Senior author.

Our data and model suggest that instantaneous social information alters both choice and confidence. After receiving the outcome, experience-derived values and observation-derived values entail comparable contributions to inform future decisions but are distinctively encoded in the vmPFC and the ACC. We further identify an interaction of two brain networks that separately process reward information and social information, and their functional coupling substantiates an RPE and a social prediction error (SPE) as teaching signals for direct learning and social learning.

RESULTS

Participants (N = 185) in groups of five performed the social influence task, of which 39 were scanned with the MRI scanner. The task design used a multiphase paradigm, enabling us to tease apart every crucial behavior under social influence (Fig. 1A). Participants began each trial with their initial choice (Choice 1) between two abstract fractals with complementary reward probabilities (70 and 30%), followed by their first postdecision bet (Bet 1, an incentivized confidence rating from 1 to 3) (*19*). After sequentially uncovering the other players' first decisions in the sequential order of participants' subjective preference (i.e., participants decided on whose choice to see in the first place and the second place, followed by the remaining two choices), participants had the opportunity to adjust their choice (Choice 2) and bet (Bet 2). The final choice and bet were then multiplied to determine the outcome on that trial (e.g., 3 × 20 = 60 cents). Participants' actual choices were communicated in real time to every other participant via intranet connections, thus maintaining a high ecological validity. The core of this paradigm was a probabilistic reversal learning (PRL) task (fig. S1B) (*20*). This PRL implementation required participants to learn and continuously relearn action-outcome associations, thus creating enough uncertainty such that group decisions were likely to be taken into account for behavioral adjustments in second decisions (before outcome delivery; referred to as instantaneous social influence) and for making future decisions on the next trial by observing others' performance (after outcome delivery; referred to as social learning) together with participants' own valuation process (referred to as direct learning). These dynamically evolving group decisions also allowed us to parametrically test the effect of group consensus, which moved beyond using only one social partner or an averaged group opinion (*2*, *12*). Although participants were able to gain full action-outcome association at the single-trial level, across trials, participants may acquire additional valuation information by observing others, given the multiple reversal nature of the PRL paradigm. In addition, participants were aware that there was neither cooperation nor competition (see Materials and Methods).

Social influence alters both action and confidence in decision-making

Human participants' choices tracked option values over probabilistic reversals (Fig. 1B). Participants indeed changed their choice and bet after observing group decisions within trials but in the opposite direction. Both the choice adjustment and the bet adjustment were modulated by a significant interaction between the relative direction of the group (with versus against) and the group consensus (2:2, 3:1, and 4:0, view of each participant; Fig. 1C). In particular, participants showed an increasing trend to switch their choice toward the group when faced with more dissenting social information, whereas they were more likely to persist when observing agreement with the group (main effect of direction: $F_{1,228}$ = 299.63, $P < 1.0 \times 10^{-15}$; main effect of consensus: $F_{2,574}$ = 131.49, $P < 1.0 \times 10^{-15}$; direction × consensus: $F_{1,574}$ = 55.82, $P < 1.0 \times 10^{-12}$; Fig. 1D). Conversely, participants tended to increase their bets as a function of the group consensus when observing confirming opinions but sustained their bets when being contradicted by the group (main effect of direction: $F_{1,734}$ = 50.95, $P < 1.0 \times 10^{-11}$; main effect of consensus: $F_{2,734}$ = 16.74, $P < 1.0 \times 10^{-7}$; direction × consensus: $F_{1,734}$ = 4.67, P = 0.031; Fig. 1E). Bet difference was also analyzed conditioned on participants' switching behavior on Choice 2, and results were in coherent with the main findings (fig. S2A).

We further verified the benefit of considering instantaneous social information for behavior adjustments. Participants' choice accuracy of the second decision was significantly higher than that of the first one (t_{185} = 3.971, $P = 1.02 \times 10^{-4}$; Fig. 1F and fig. S2B), and participants' second bet was significantly larger than their first one (t_{185} = 2.665, P = 0.0084; Fig. 1G and fig. S2C). These results suggested that, in the case of behavioral adjustments, despite that participants were often confronted with conflicting group decisions, considering social information in fact facilitated learning. Notably, these behavioral adjustments were not likely due to perceptual conflict, in which participants would have randomly made switches, hence no learning enhancement. Notably, no such benefit of adjustment was observed in the nonsocial control experiment, where participants (N = 36; note S1) were performing this task with intelligent computer agents (fig. S1, A, C to F). Note that although we did not intentionally manipulate the amount of dissenting social information in the main experiment (given the real-time property), it was nonetheless randomly distributed (Wald-Wolfowitz runs test, all $P > 0.05$). Moreover, neither the amount of dissenting social information nor participants' choice switching behavior was related to the time of reversal or the lapse error indicated by our winning model (see Materials and Methods and fig. S2, E and F).

Furthermore, we assessed the learning benefit of considering social information between trials. We found that the accuracy of Choice 1 on the current trial was modulated by a significant three-way interaction among choice adjustment on the previous trial [Choice 2 switch versus stay (SwSt)], the relative direction of the group (with versus against), and the group consensus (2:2, 3:1, and 4:0) (main effect of Choice 2 type: $F_{1,1604}$ = 14.52, $P = 1.44 \times 10^{-4}$; Choice 2 type × direction: $F_{1,1604}$ = 79.12, $P = <1.0 \times 10^{-15}$; Choice 2 type × consensus: $F_{1,1604}$ = 6.27, P = 0.0019; Choice 2 type × direction × consensus: $F_{1,1604}$ = 16.89, $P < 1.0 \times 10^{-4}$; no other effects were significant, all $P > 0.05$; Fig. 1H). In particular, the accuracy of the current choice was improved either when more opposing choices were observed and participantss switched between trials or when confirming choices were observed and participants retained their choice between trials. A significant three-way interaction was also observed for the magnitude of Bet 1 on the current trial, suggesting that participants were overall more confident following others' choices on the previous trial and this held for both stay or switch decisions (main effect of Choice 2 type: $F_{1,1597}$ = 136.34, $P < 1.0 \times 10^{-15}$; Choice 2 type × direction: $F_{1,1597}$ = 56.92, $P = <1.0 \times 10^{-13}$; Choice 2 type × consensus: $F_{1,1597}$ = 8.96, $P = 1.35 \times 10^{-4}$; Choice 2 type × direction × consensus: $F_{1,1597}$ = 5.98, P = 0.0146; no other effects were significant, all $P > 0.05$; Fig. 1I). Notably, in the nonsocial control experiment, although participants had a similar choice pattern to that of the main experiment (fig. S1H), no effect of confidence was observed

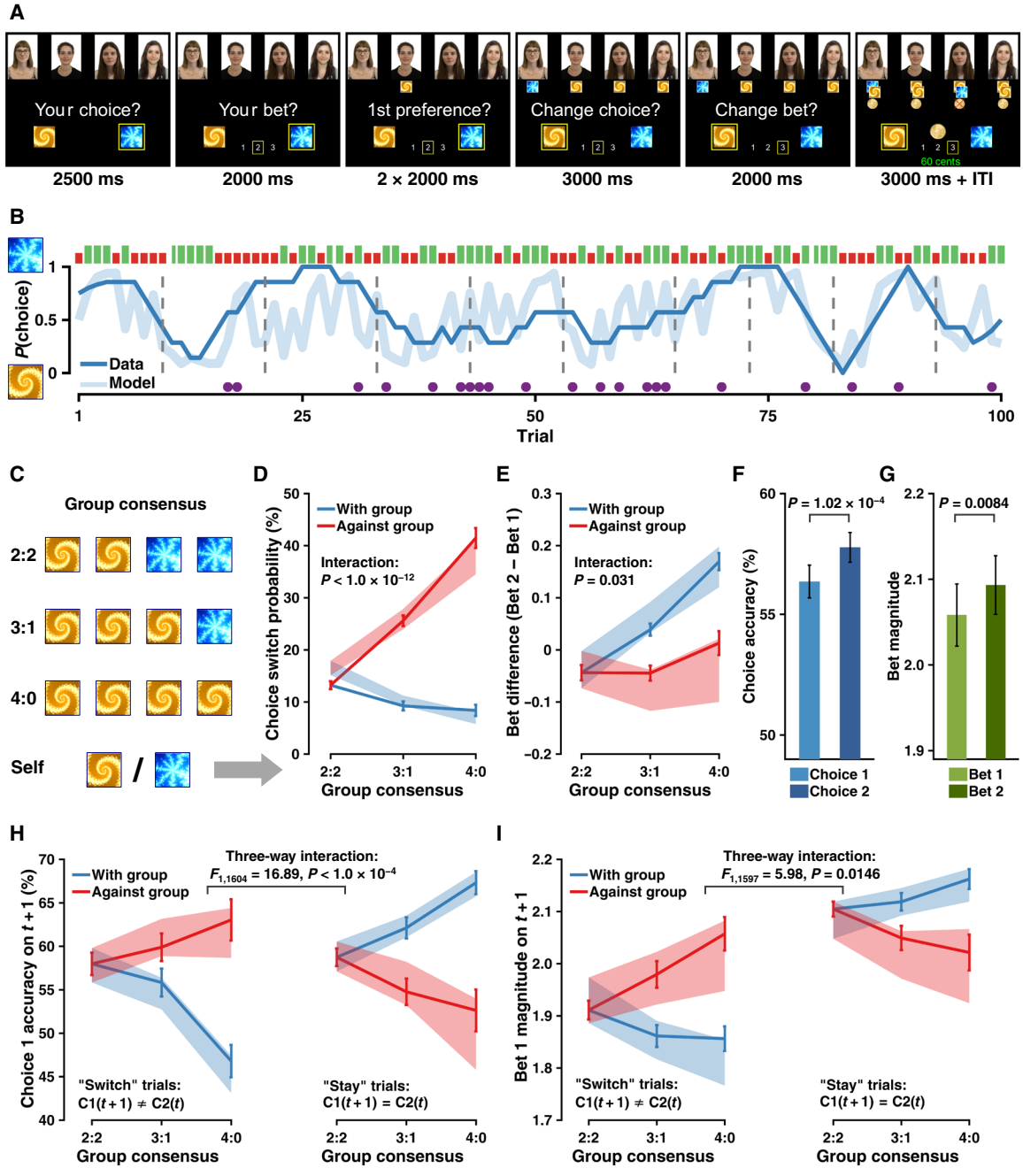

Fig. 1. Experimental task and behavioral results. (**A**) Task design. Participants ($N = 185$) made an initial choice and bet (Choice 1 and Bet 1), and after observing the other four coplayers' initial choices, they were asked to make adjustments (Choice 2 and Bet 2), followed by the outcome. (**B**) Task dynamic. Trial-by-trial behavior for an example participant. Blue curves, seven-trial running averages (dark) and predicted choice probabilities from the winning model M6b (light). Green (long) and red (short) bars, rewarded and unrewarded trials; purple circles, switches on Choice 2; dashed vertical lines, reversals every 8 to 12 trials. (**C**) Illustration of group consensus (view from each participant). (**D**) Social influence on within-trial choice adjustments. Choice switch probability as a function of group consensus [as (C)] and relative direction (with versus against) of the group. Solid lines indicate actual data [means ± within-subject SE (SEM)]. Shaded error bars represent 95% highest density interval (HDI) of mean effects predicted by M6b (i.e., posterior predictive checks). (**E**) Social influence on within-trial bet adjustments. Bet difference as a function of group consensus and direction of the group. Format is as in (D). (**F** and **G**) Enhanced Choice 2 and Bet 2 performance after adjustment. (**H**) Social influence on between-trial choice accuracy. Choice 1 accuracy on the current trial as a function of choice adjustment on the previous trial (Choice 2 type: SwSt), direction of the group, and group consensus. Format is as in (D). (**I**) Social influence on between-trial bet magnitude. Bet 1 magnitude on the current trial as a function of choice adjustment on the previous trial, relative direction of the group, and group consensus. Format is as in (D).

between trials (fig. S1I). These results may indicate that behaviors in main experiment manifested goal emulation, whereas behaviors in the nonsocial control experiments reflected choice imitation (*17*). Together, our behavioral results demonstrated that social information altered individuals' choice and confidence, which accounted for facilitated learning after behavioral adjustment both

within trials and between trials, and this benefit could not be explained by perceptual mismatch and may be specific only when interacting with human partners.

Computational mechanisms of integrated valuation between direct learning and social learning

Using computational modeling, we aimed to formally quantify latent mechanisms that underlay the learning processes in our task on a trial-by-trial basis. Different from existing RL models on social learning of a confederate's advice (*12*), our model accommodates multiple players and is able to simultaneously estimate all participants' behaviors (two choices and two bets). Our efforts to construct the winning model (Fig. 2A) were guided by two design principles: (i) separating individual's own valuation updated via direct learning from vicarious valuation updated via social learning and (ii) distinguishing instantaneous social influence before outcome delivery from social learning in which action-outcome associations were observed from the others. These design principles tied closely with our multiple task phases, representing a computationally plausible information flow.

On each trial, the option value of Choice 1 (A or B) was modeled as a linear combination between values from direct learning (V_{self}) and values from social learning (V_{other})

$$\mathbb{V}_t = \beta_{vself}\,\mathbb{V}_{self,t} + \beta_{vother}\,\mathbb{V}_{other,t} \quad (1)$$

where

$$\begin{aligned}\mathbb{V}_{self,t} &= [V_{self,t}(A), V_{self,t}(B)]\\ \mathbb{V}_{other,t} &= [V_{other,t}(A), V_{other,t}(B)]\end{aligned} \quad (2)$$

After participants found the other coplayers' first choices, their Choice 2 (switch or stay) was modeled as a function of two counteracting influences: (i) the preference-weighted group dissension ($w.N_{against}$) representing the instantaneous social influence and (ii) the difference between participants' action values of Choice 1 ($V_{chosen,C1,t} - V_{unchosen,C1,t}$) representing the distinctiveness of the current value estimates.

Last, when all outcomes were delivered, both V_{self} and V_{other} were updated. Notably, V_{self} was updated using the fictitious Rescorla-Wagner RL model (*20*, *21*) (Fig. 2B), whereas V_{other} was updated through tracking an exponentially decayed and preference-weighted all four other coplayers' cumulative reward histories (i.e., their performance in the recent past; Fig. 2C). Note that our construction of V_{other} was in close accordance with previous evidence that suggested a discounted outcome history contributing to animals' valuation processes (*22*) and that the construction of V_{other} depicted social learning by simulating a vicarious valuation process by observing others (*4*, *13*, *17*, *23*). The social learning here was weighted by social preference ($w_{s,t}$) that reflected credibility assignment based on the social partners' performance (*12*, *16*). V_{other} did not contribute to the learning performance in the nonsocial control task despite similar behavioral adjustment patterns compared to the main study, suggesting the uniqueness of social learning in social contexts (fig. S1G). Together, all the above properties granted the social feature of V_{other} and demonstrated its distinct contribution in addition to V_{self} (Fig. 2D).

We tested the winning model against alternative computational hypotheses under the hierarchical Bayesian framework (Table 1) (*24*). We further verified our winning model using two rigorous validation approaches. First, we carried out a parameter recovery analysis to assure that all parameters could be accurately and selectively identified (note S3 and fig. S3). Second, as model comparison provided relative model performance, we noted the importance to perform posterior predictive checks, and we found that the posterior prediction well captured key behavioral patterns (Fig. 1, D, E, H, and I, and fig. S2A).

Parameter estimation results (Fig. 2, E to H) suggested that the extent to which participants learned from themselves and from the others was on average comparable {$\beta(V_{self}) = 0.84$, 95% highest density interval (HDI): [0.67, 1.01]; $\beta(V_{other}) = 0.78$, 95% HDI: [0.59, 0.97]}, suggesting that value signals computed from direct learning and social learning were jointly used to guide future decisions. These results were corroborated by a multiple regression where both $\beta(V_{self})$ (effect = 0.033, $P < 1.0 \times 10^{-5}$) and $\beta(V_{other})$ (effect = 0.024, $P = 0.0014$) significantly predicted the accuracy of Choice 1. Possible modulation of Bet 1 on $\beta(V_{self})$ and $\beta(V_{other})$ was also assessed (note S4 and fig. S2D). Furthermore, parameters related to instantaneous social information were well capable of predicting individual differences of participants' behavioral adjustment: If the model-derived signal was in high accordance with the corresponding pattern of behavioral adjustment, then we ought to anticipate a strong association between them. We observed a positive correlation between $\beta(w.N_{against})$ and slopes of choice switch probabilities in the against condition ($r = 0.64$, $P < 1.0 \times 10^{-21}$; Fig. 2I; slopes computed from Fig. 1D). Similarly, we observed a positive correlation between $\beta(w.N_{with})$ and slopes derived from bet differences in the "with" condition ($r = 0.33$, $P < 1.0 \times 10^{-5}$; Fig. 2J; slopes computed from Fig. 1E). Together, our computational modeling analyses suggested that participants learned both from their direct valuation process and from vicarious valuation experience, and values from direct learning and social learning jointly contributed to the decision process. Moreover, participants' behavioral adjustments were predicted by the counteracting effects between their initial valuation and the instantaneous social information. Next, once we had uncovered those latent variables of the decision processes underlying the social influence task, we were able to test how they were computed and implemented at the neural level using model-based fMRI (*25*).

Neural substrates of dissociable value signals from direct learning and social learning

The first part of our model-based fMRI analyses focused on how distinctive decision variables (Fig. 3A) were represented in the brain [general linear model 1 (GLM 1)]. We aimed to test the hypothesis that distinct and dissociable brain regions were recruited to implement direct learning and social learning signals (i.e., component value) (*3*). We observed that the vmPFC [4, 46, −14; see table S4 for all Montreal Neurological Institute (MNI) coordinates and multiple comparisons correction methods] activity was positively scaled with V_{self} and the ACC (2, 10, 36) activity was positively scaled with V_{other} (Fig. 3B). To test whether the two value signals were distinctively associated with vmPFC and ACC, we used a double-dissociation approach, and we found that V_{self} was exclusively encoded in the vmPFC ($\beta = 0.1458$, $P < 1.0 \times 10^{-5}$; Fig. 3E, red) but not in the ACC ($\beta = 0.0128$, $P = 0.4394$; Fig. 3D, red), whereas V_{other} was exclusively represented in the ACC ($\beta = 0.1560$, $P < 1.0 \times 10^{-5}$; Fig. 3D, blue) but not in the vmPFC ($\beta = 0.0011$, $P = 0.9478$; Fig. 3E, blue). Computationally, these two sources of value signals needed to be integrated to make future decisions

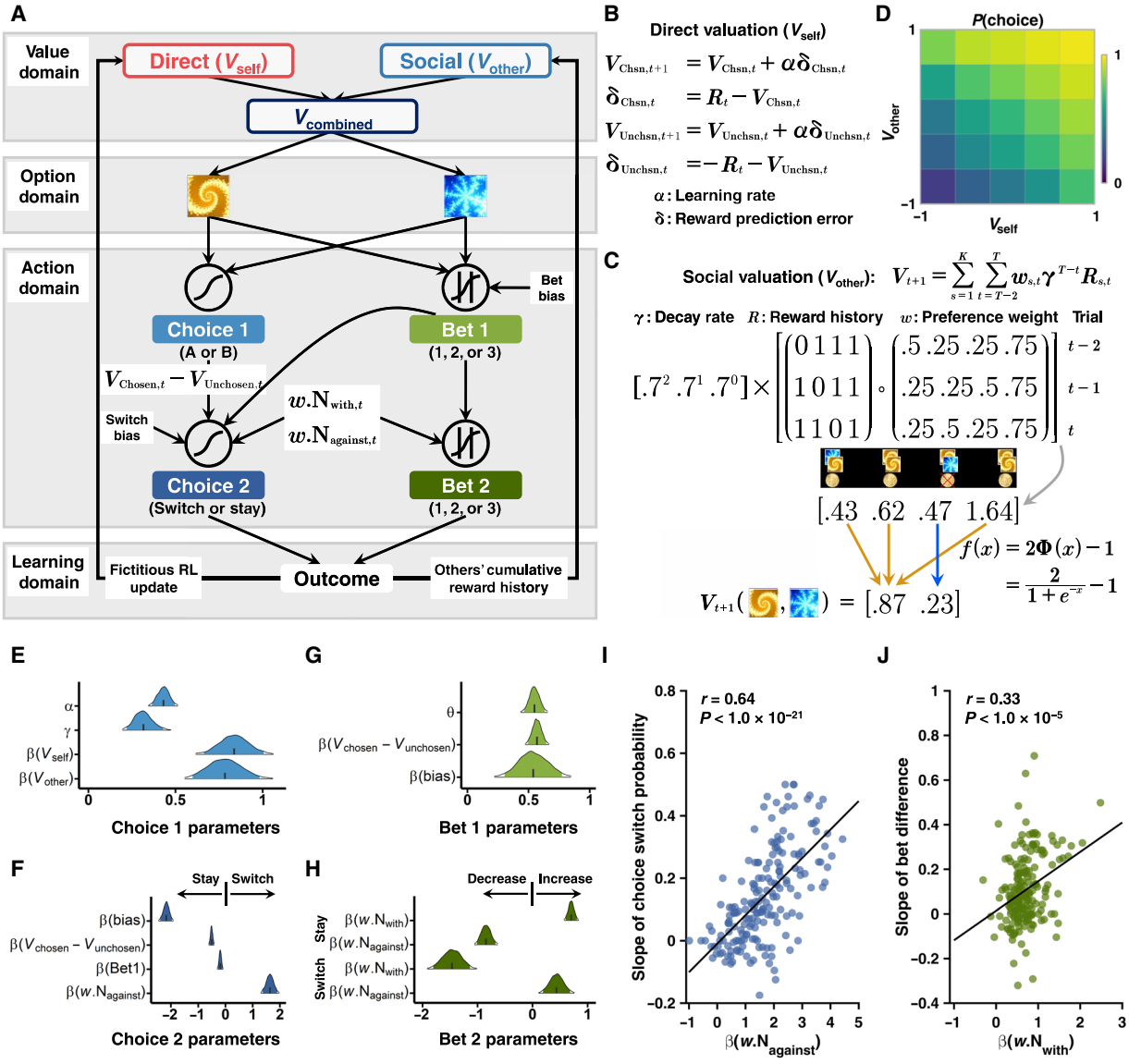

Fig. 2. Computational model and its relation to behavior. (**A**) Schematic representation of the winning model (M6b). Participants' initial behaviors (Choice 1 and Bet 1) were accounted for by value signals updated from direct learning (V_{self}) and social learning (V_{other}); behavioral adjustments (Choice 2 and Bet 2) were ascribed to initial valuation ($V_{chosen,t} - V_{unchosen,t}$) and preference-weighted instantaneous social information ($w.N_{with,t}$ and $w.N_{against,t}$). (**B**) Computation of V_{self}. V_{self} was updated via the fictitious RL model, where values of both choice options were updated. (**C**) Computation of V_{other}. V_{other} was updated through tracking other coplayers' cumulative reward histories in the last three trials ($t-2$ to t), weighted by preference and a decay rate, and further normalized to lie between −1 and 1. The ° sign indicates the Hadamard product (i.e., element-wise product). (**D**) Contribution of V_{self} and V_{other} to action probability of Choice 1. Both V_{self} and V_{other} spanned within their range (−1 to 1), and they jointly contributed to *P*(Choice 1). (**E** to **H**) Model parameters of M6b. Posterior density for parameters related to Choice 1 (E), Choice 2 (F), Bet 1 (G), and Bet 2 (H). Short vertical bars indicate the posterior mean. Shaded areas depict 95% HDI. (**I** and **J**) Relationship between model parameters and behavioral results across participants. (I) Relationship between the effect of dissenting social information β($w.N_{against}$) and the susceptibility to social influence (i.e., slope of switch probability calculated from Fig. 1D). (J) Relationship between the effect of confirming social information β($w.N_{with}$) and the extent of bet difference (i.e., slope of bet difference calculated from Fig. 1E).

(i.e., integrated value) (*3*). We reasoned that if a region was implementing the integrated value, then it must show functional connectivities with regions tracking each of the value signals (i.e., vmPFC and ACC). Using a physio-physiological interaction (PhiPI) analysis, we found that the medial prefrontal cortex (mPFC; 10, 40, 10) covaried with both the vmPFC and the ACC (fig. S6A).

Besides the value signals, the RPE signal was firmly associated with activities in the bilateral NAcc (Fig. 3C; left: −10, 8, −10; right: 12, 10, −12). Furthermore, a closer look at the two theoretical subcomponents of RPE was necessary to assess its neural substrates (*12*, *26*). Specifically, according to the specification of RPE (Fig. 2B), to qualify as a region encoding the RPE signal, activities in the NAcc ought to covary positively with the actual outcome (e.g., reward) and negatively with the expectation (e.g., value). This property thus provides a common framework to test the neural correlates of any error-like signal. Under this framework, we indeed found that

Table 1. Candidate computational models, model comparison, and winning model's parameters. SL, social learning; # Par., number of free parameters at the individual level; ΔLOOIC, leave-one-out information criterion relative to the winning model (lower LOOIC value indicates better out-of-sample predictive accuracy); weight, model weight calculated with Bayesian model averaging using Bayesian bootstrap (higher model weight value indicates higher probability of the candidate model to have generated the observed data). M6b (in bold) is the winning model.

Class	Model	Description	# Par.	ΔLOOIC	Weight
Nonsocial models	M1a	Simple Rescorla-Wagner RL	9	3614	0
	M1b	Fictitious RL	9	2369	0
	M1c	Pearce-Hall	11	8540	0
Social models: instantaneous social influence	M2a	M1a + instantaneous social influence	9	1721	0
	M2b	M1b + instantaneous social influence	9	725	0
	M2c	M1c + instantaneous social influence	11	2715	0
Social models: instantaneous social influence and social learning	M3	M2b + SL (others' RL update)	15	535	0.002
	M4	M2b + SL (others' action preference)	13	745	0
	M5	M2b + SL (others' current reward)	13	411	0
	M6a	M2b + SL (others' cumulative reward)	14	164	0
	M6b	M2b + SL (others' cumulative reward) + Bet 1	15	0	0.998

M6b choice-related parameters (mean and 95% HDI)

Choice 1		Choice 2	
α	0.43 [0.37, 0.50]	β(bias)	−2.17 [−2.39, −1.96]
γ	0.32 [0.22, 0.43]	$\beta(V_{chosen} - V_{unchosen})$	−0.51 [−0.59, −0.43]
$\beta(V_{self})$	0.84 [0.66, 1.01]	β(Bet 1)	−0.19 [−0.28, −0.11]
$\beta(V_{other})$	0.78 [0.60, 0.98]	$\beta(w.N_{against})$	1.63 [1.37, 1.90]

M6b bet-related parameters (mean and 95% HDI)

Bet 1		Bet 2	
θ	0.55 [0.46, 0.63]	$\beta(w.N_{with}\|stay)$	0.70 [0.60, 0.80]
$\beta(V_{chosen} - V_{unchosen})$	0.57 [0.50, 0.64]	$\beta(w.N_{against}\|stay)$	−0.85 [−1.01, −0.68]
β(bias)	0.54 [0.30, 0.77]	$\beta(w.N_{with}\|switch)$	−1.47 [−1.80, −1.14]
		$\beta(w.N_{against}\|switch)$	0.44 [0.19, 0.68]

activities in the NAcc showed a positive effect of participants' actual reward (i.e., R; $\beta = 0.2298$, $P < 1.0 \times 10^{-5}$) and a negative effect of their valuation (i.e., V_{self}; $\beta = -0.0327$, $P = 0.021$; Fig. 3F), justifying that NAcc was encoding the RPE signal instead of the outcome valence. Variables related to participants' bet did not yield significant clusters.

Neural correlates of dissenting social information and behavioral adjustment

We next turned to disentangle the neural substrates of the instantaneous social influence (GLM 1) and the subsequent behavioral adjustment (GLM 2). Since we have validated enhanced learning after considering instantaneous social information (Fig. 1, F and G), we reasoned that participants might process other coplayers' intentions relative to their own first decision to make subsequent adjustments and this might be related to the mentalizing network (*17*). On the basis of this reasoning, we assessed the parametric modulation of preference-weighted dissenting social information ($w.N_{against}$) and found that activities in the TPJ (left: −48, −62, 30; right: 50, −60, 34), among other regions, were positively correlated with the dissenting social information (fig. S4). Furthermore, the resulting choice adjustment (i.e., switch > stay) covaried with activity in bilateral dorsolateral prefrontal cortex (dlPFC) (fig. S5, A and D; left: −32, 48, 16; right: 26, 42, 32), commonly associated with executive control and behavioral flexibility across species (*20*, *27*). By contrast,

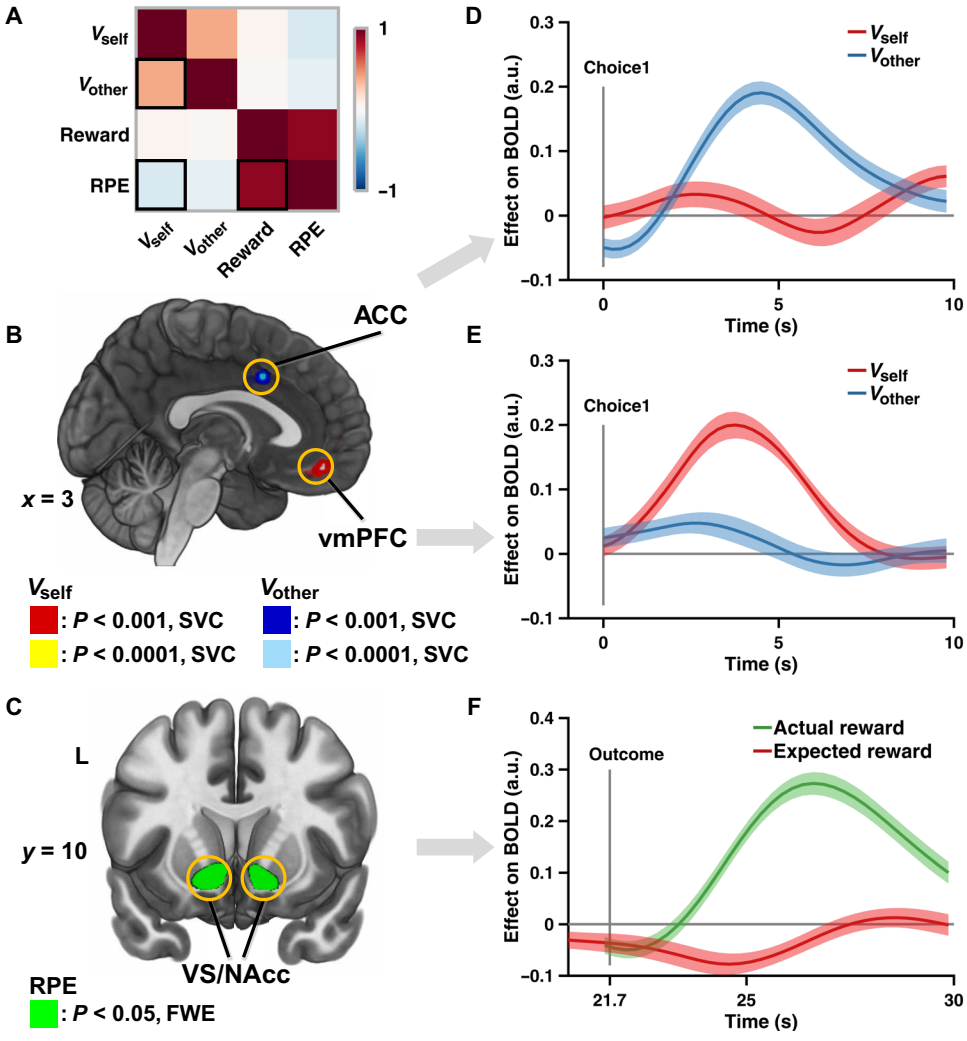

Fig. 3. Neural substrates of dissociable value signals and RPE. (**A**) Correlation matrix of value-related decision variables derived from M6b. (**B**) Neural representation of value signals. V_{self} and V_{other} were encoded in the vmPFC (red/yellow) and the ACC (blue/light blue), respectively. Sagittal slice at $x = 3$. Display thresholded at $P < 0.001$ and $P < 0.0001$, small volume–corrected (SVC); actual results were threshold-free cluster enhancement (TFCE) SVC at $P < 0.05$. (**C**) Neural representation of RPE. RPE was encoded in the VS/NAcc. Coronal slice at $y = 10$. Display thresholded at $P < 0.05$, family-wise error (FWE) corrected; actual results were TFCE whole-brain FWE corrected at $P < 0.05$. (**D** and **E**) Region of interest (ROI) time series analyses of vmPFC and ACC, demonstrating a double dissociation of the neural signatures of value signals. (D) Blood oxygen level–dependent (BOLD) signal of ACC was only positively correlated with V_{other} ($\beta = 0.1560$, $P < 1.0 \times 10^{-5}$, permutation test; blue line), but not with V_{self} ($\beta = 0.0011$, $P = 0.9478$, permutation test; red line), whereas (E) BOLD signal of vmPFC was only positively correlated with V_{self} ($\beta = 0.1458$, $P < 1.0 \times 10^{-5}$, permutation test; red line) but not with V_{other} ($\beta = 0.0128$, $P = 0.4394$, permutation test; blue line). Lines and shaded areas show means ± SEM of β weights across participants. (**F**) ROI time series analyses of VS/NAcc, showing its sensitivity to both components of RPE (i.e., actual reward R and expected reward V_{self}). BOLD signal of VS/NAcc was positively correlated with actual reward ($\beta = 0.2298$, $P < 1.0 \times 10^{-5}$, permutation test; green line) and negatively correlated with expected reward ($\beta = -0.0327$, $P = 0.021$, permutation test; red line). Format is as in (D). a.u., arbitrary units.

the vmPFC was more active during stay trials (i.e., stay > switch), reminiscent of its representation of one's own valuation (fig. S5, C and F). Hence, these findings were not likely due to learning of the task structure but rather were genuinely attributed to dissenting social information and choice adjustment, respectively.

A network between brain's reward circuits and social circuits

Above, we demonstrated how key decision variables related to value and reward processing and social information processing were implemented at distinct nodes at the neural level. In the next step, we sought to establish how these network nodes were functionally connected to bring about socially induced behavioral change and to uncover additional latent computational signals that would otherwise be undetectable by conventional GLMs.

Using a psychophysiological interaction (PPI) (*28*), we investigated how behavioral change at Choice 2 was associated with the functional coupling between rTPJ that processed instantaneous social information and other brain regions. This analysis identified enhanced connectivity between left putamen (lPut; Fig. 4, A to C; −20, 12, −4) and rTPJ as a function of choice adjustment. Closer investigations into the computational role of lPut revealed that it did not correlate with both subcomponents of the RPE (fig. S6C). Instead, as the choice adjustment resulted from processing social information, we reasoned that lPut might encode an SPE at the time

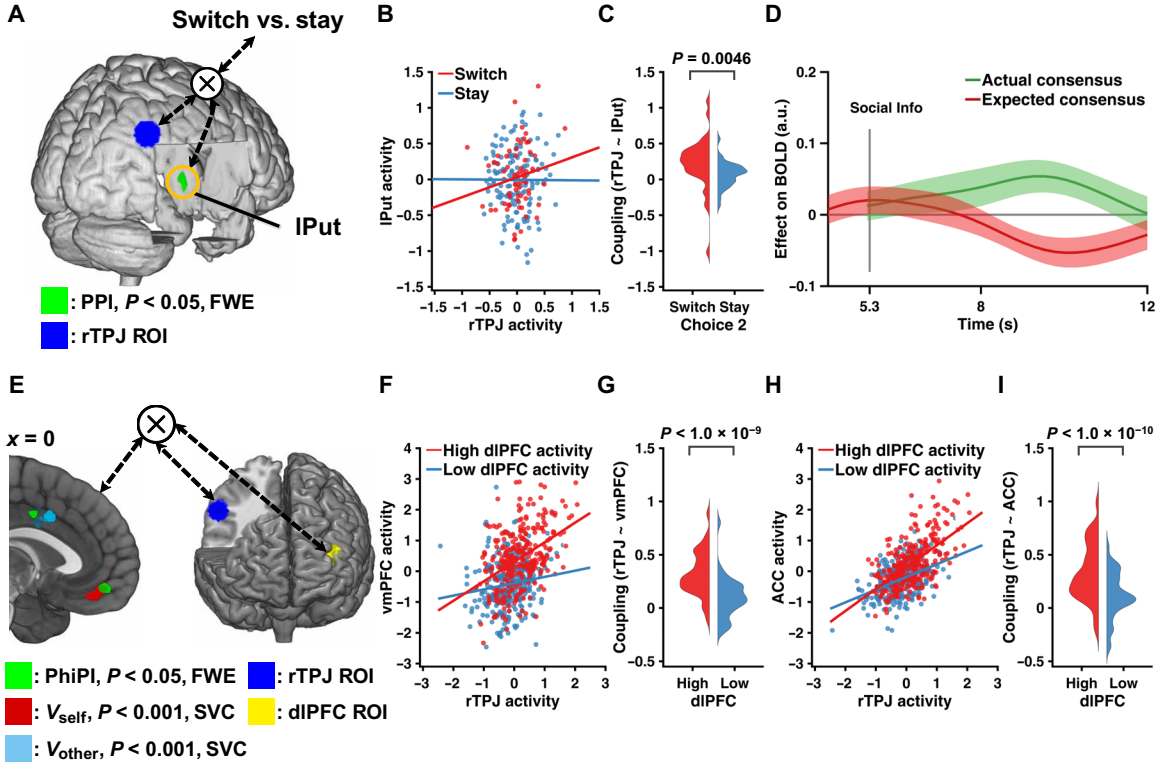

Fig. 4. Functional connectivity between reward-related regions and social-related regions. (**A**) Increased functional connectivity between the lPut (green) and the seed region rTPJ (blue) as a function of choice adjustment (SwSt). Display thresholded at $P < 0.05$, FWE corrected. Actual results were TFCE whole-brain FWE corrected at $P < 0.05$. We seeded at the rTPJ because its suprathreshold cluster was larger than the left TPJ (lTPJ) (table S4). Using lTPJ as the seed yielded similar yet slightly weaker results. (**B**) Correlation of activity in seed and target regions for both switch and stay trials in an example participant. (**C**) Kernel density estimation of coupling strength across all participants for switch and stay trials. (**D**) ROI time series analyses of the lPut, exhibiting an SPE signal: BOLD signal of lPut was positively correlated with the actual consensus ($\beta = 0.0363$, $P = 0.0438$, permutation test; green line) and negatively correlated with the expected consensus ($\beta = -0.0409$, $P = 0.0123$, permutation test; red line). Format is as in Fig. 3F. (**E**) PhiPI between social-related regions and reward-related regions. The rTPJ seed (blue) and the left dlPFC seed (yellow) elicited connectivity activations (target regions) in the vmPFC and the pMFC (both in green), which partially overlapped with neural correlates of value signals in vmPFC and ACC, as in Fig. 3B. Sagittal slice at $x = 0$. Display thresholded at $P < 0.05$, FWE corrected; actual results were TFCE whole-brain FWE corrected at $P < 0.05$. (**F** to **I**) Correlation plots of seed and target regions for both high and low dlPFC activities in an example subject (F and H) and kernel density estimation of seed-target coupling strengths across all participants for high and low dlPFC activities (G and I).

of observing social information, delineating the difference between the actual consensus and the expected consensus of the group. Specifically, the expected consensus was approximated by the difference in participants' vicarious valuation ($V_{other,chosen,t} - V_{other,unchosen,t}$), on the basis that knowing how the others value specific options helps individuals model the others' future behaviors (*23*, *29*) (e.g., when $V_{other,chosen,t} - V_{other,unchosen,t}$ was large, participants were relatively sure about option values learned from the others, therefore anticipating more coherent group choices). Following this reasoning, we conducted a similar time series analysis as we did for the RPE, and we found that activity in the lPut was indeed positively correlated with the actual consensus ($\beta = 0.0363$, $P = 0.0438$) and negatively correlated with the expected consensus ($\beta = -0.0409$, $P = 0.0123$; Fig. 4D). This pattern suggested that lPut was effectively encoding a hitherto uncharacterized SPE rather than an RPE (fig. S6B). Together, these analyses demonstrated that the functional coupling between neural representations of social information and SPE was enhanced, when this social information was leading to a behavioral change.

In the last step, using a PhiPI, we investigated how neural substrates of switching at Choice 2 in the left dlPFC were accompanied by the functional coupling of rTPJ and other brain regions. This analysis revealed that rTPJ covaried with both vmPFC (0, 48, −12) and ACC (0, 0, 40), scaled by the activation level of dlPFC (Fig. 4, E to I). Notably, these target regions overlapped with regions that represented two value signals in vmPFC and ACC that we reported earlier (c.f. Fig. 3B). Collectively, our functional connectivity analyses suggested the interplay of brain regions representing social information and the propensity for behavioral change led to the neural activities of values signals in the vmPFC and ACC, which are updated via both direct learning and social learning.

DISCUSSION

Social influence is a powerful modulator of individual choices, yet how social influence and subsequent social learning interact with direct learning in a probabilistic environment is poorly understood. Here, we bridge this gap with a multiplayer social decision-making paradigm in real time that allowed us to dissociate between experience-driven valuation and observation-driven valuation. In a comprehensive neurocomputational approach, we are not only able to identify a network of brain regions that represents and integrates

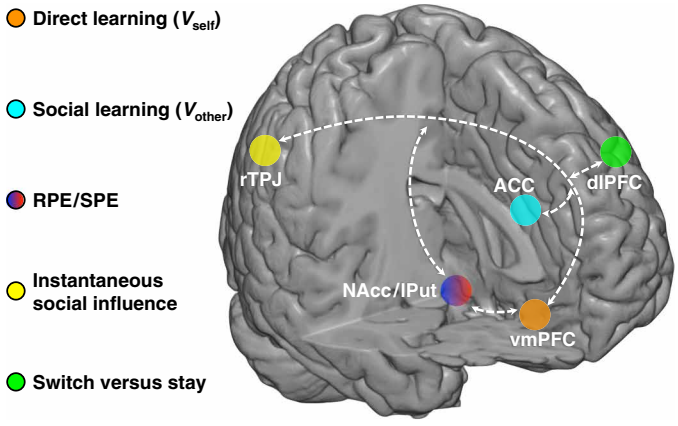

Fig. 5. Schematic illustration of the brain network supporting social influence in decision-making as uncovered in this study (for details, see main text).

social information in learning but also characterize the computational role of each node in this network in detail (Fig. 5), suggesting the following process model: Individuals' own decision is guided by a combination of value signals from direct learning (V_{self}) represented in the vmPFC (Fig. 3, B and E) and from social learning (V_{other}) represented in a section of the ACC (Fig. 3, B and D). The instantaneous social information reflected by decisions from others is encoded with respect to one's own choice in the rTPJ (fig. S4), an area linked, but not limited to representations of social information and social agents in a variety of tasks (*16*, *17*). rTPJ is also related to Theory of Mind (*30*) and other integrative computations including multisensory integration (*31*). Moreover, dissenting social information gives rise to a hitherto uncharacterized SPE (difference between actual and expected consensus of the group) represented in the putamen (Fig. 4D), unlike the more medial NAcc, which exhibits the neural signature of a classic RPE (Fig. 3, C and F) (*6*). Notably, the interplay of putamen and rTPJ modulates behavioral change toward the group decision (Fig. 4, A to C) in combination with its neural representation of choice switching in the dlPFC (Fig. 4, E to I). These connected neural activations functionally couple with the valuation of direct learning in the vmPFC (V_{self}) and social learning in the ACC (V_{other}), thus closing the loop of decision-related computations in social contexts.

Our result that direct valuation is encoded in the vmPFC is firmly in line with previous evidence from learning and decision-making in nonsocial contexts (*9*) and demonstrates the role of vmPFC in experiential learning into a social context. In addition to individuals' own value update, we further show that the ACC encodes value signals updated from social learning, which is aligned with previous studies that have implicated the role of ACC in tracking the volatility of social information (*12*) and vicarious experience (*32*). In particular, given that social learning in the current study is represented by the preference-weighted cumulative reward histories of the others, the dynamics of how well the others were performing in the recent past somewhat reflect their volatility in the same learning environment (*12*). Moreover, this distinct neural coding of direct values and vicarious values in the current study fundamentally differs from previous studies on social decision-making. While previous studies have found evidence for a role of vmPFC and ACC in encoding self-oriented and other-oriented information (*14*, *33*), those signals were invoked when participants were explicitly requested to alternately make decisions for themselves or for others. Crucially in the present study, because direct learning and social learning coexisted in the probabilistic environment and no overt instruction was given to differently track oneself and the others, we argue that these two forms of learning processes are implemented in parallel, and our winning model indicates that the extent to which individuals rely on their own and the others is effectively comparable. Thus, the neurocomputational mechanisms being revealed here are very distinct from those that have been reported previously. Taken collectively, these results demonstrate coexisting, yet distinct value computations in the vmPFC and the ACC for direct learning and social learning, respectively, and are in support of the social valuation–specific schema (*23*).

Our functional connectivity analyses revealed that the mPFC covaried with activations in both vmPFC and ACC. According to a recent meta-analysis (*9*), this region is particularly engaged during the decision stage when individuals are representing options and selecting actions, especially in value-based and goal-directed decision-making (*34*). Hence, it suggests that beyond the dissociable neural underpinnings, the direct value and vicarious value are further combined to make subsequent decisions.

Furthermore, we replicated previous evidence that NAcc is associated with the RPE computation instead of mere outcome valence (*12*, *26*). That is, if a brain region encodes the RPE, then its activity should be positively correlated with the actual outcome and negatively correlated with the expected outcome. Beyond reassuring the RPE signal encoded in the NAcc, the corresponding time series analysis serves as a verification framework for testing neural correlates of any error-like signals. Hence, our connectivity results seeded at the rTPJ identified a hitherto uncharacterized SPE, the difference between actual and expected social outcome, that is encoded in a section of the putamen. This suggests that the SPE signal may trigger a recomputation of expected values and give rise to the subsequent behavioral adjustment. We nonetheless acknowledge that the connectivity analyses here assess correlation rather than directionality and establishing the casual account using brain stimulations (*35*) or pharmacological manipulations (*36*) would be a promising avenue for future work. Albeit this methodological consideration, these functional connectivity results concur with previous evidence that the rTPJ has functional links with the brain's reward network, of which the striatal region is a central hub (*37*).

It is perhaps unexpected and interesting that we did not find significant neural correlates with postdecision confidence (i.e., "bet"). This might be due to the fact that decision cues in our current design (i.e., Choice 1, Bet 1, Choice 2, and Bet 2) were not presented far apart in time, such that even carefully specified GLMs were not able to capture the variance related to bets. Bets in the current design were closely tied to the corresponding choice valuation. That is, when individuals were sure that one option would lead to a reward, they tended to place a high bet. This relationship was well reflected in our winning model and related model parameters (Fig. 2G): Bet magnitude was positively correlated with value signals, thus inevitably resulting in colinear regressors and diminishing the statistical power when assessing its neural correlates. In addition, individuals' response time might be needed to dissociate confidence from their valuation (*18*). These caveats aside, our results nonetheless shed light on the change in confidence after incorporating social information in decision-making, which goes beyond evidence from previous studies that neither directly addressed the difference in confidence before and after exposing the

social information nor examined the interface between choice and confidence (*3*).

Note also that the model space in the current study is not exhaustive. In particular, we did not test Bayesian models that would track more complex task dynamics (*38*, *39*), as this class of models may not give advantages in our task environment (*40*). The complexity of our task structure, with making four sets of choices and bets and observing two sets of actions as well as the action-outcome associations from four other coplayers, made the construction of explicit representation prescribed by Bayesian models rather challenging. In addition, it is so far still unanswered whether RL-like models or Bayesian models provide a more veridical description of how humans make decisions under uncertainty (*41*). Regardless of this debate, our fictitious RL model implemented for direct learning is reconciled with previous findings showing its success in reversal learning tasks in both humans (*20*) and nonhuman primates (*15*, *27*).

In summary, our results provide behavioral, computational, and neural evidence for dissociable representations of direct valuation learned from own experience and vicarious valuation learn from observations of social partners. Moreover, these findings suggest a network of distinct, yet interacting brain regions substantiating crucial computational variables that underlie these two forms of learning. Such a network is in a prime position to process decisions of the sorts mentioned at the beginning, where—as in the example of a lunch order—we have to balance our own experienced-based reward expectations with the expectations of congruency with others and use the resulting error signals to flexibly adapt our choice behavior in social contexts.

MATERIALS AND METHODS

Participants

Forty-one groups of five healthy, right-handed participants were invited to participate in the main experiment. No one had any history of neurological and psychiatric diseases nor current medication except contraceptives or any MR-incompatible foreign object in the body. To avoid gender bias, each group consisted of only same-gender participants. To avoid familiarity bias, we explicitly specified in the recruitment that if friends were signing up, then they should sign up for different sessions. Forty-one of 205 participants (i.e., one of each group) were scanned with fMRI while undergoing the experimental task. The remaining 164 participants were engaged in the same task via intranet connections while being seated in the adjacent behavioral testing room outside the scanner. Twenty of 205 participants who had only switched once or had no switch at all were excluded, including two fMRI participants. This was to ensure that the analysis was not biased by these nonresponders. The final sample consisted of 185 participants (95 females; mean age, 25.56 ± 3.98 years; age range, 18 to 37 years), and among them, 39 participants belonged to the fMRI group (20 females; mean age, 25.59 ± 3.51 years; age range, 20 to 37 years).

All participants in both studies gave informed written consent before the experiment. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Medical Association of Hamburg (PV3661).

Experimental design

Underlying PRL paradigm

The core of our social influence task was a PRL task. In our two-alternative forced choice PRL (fig. S1B), each choice option was associated with a particular reward probability (i.e., 70 and 30%). After a variable length of trials (length randomly sampled from a uniform distribution between 8 and 12 trials), the reward contingencies reversed, such that individuals who were undergoing this task needed to readapt to the new reward contingencies so as to maximize their outcome. Given that there was always a "correct" option, which led to more reward than punishment, alongside an "incorrect" option, which caused otherwise, a higher-order anti-correlation structure thus existed to represent the underlying reward dynamics. This task specification also laid the foundation for our use of fictitious RL model with counterfactual updating (*15*, *20*, *27*).

We used the PRL task rather than tasks with constant reward probability (e.g., always 70%) because the PRL task structure required participants to continuously pay attention to the reward contingency, to adapt to the potentially new state of the reward structure, and to ignore the (rare) probabilistic punishment from the correct option. As a result, the PRL task assured constant learning throughout the entire experiment: Choice accuracy reduced after reversal took place but soon reinstated (fig. S2, B and C). One of our early pilot studies used a fixed reward probability. There, participants quickly learned the reward contingency and neglected the social information; thus, in this setup, we could not tease apart the contributions between reward-based influence and socially based influence.

Breakdown of the social influence task (main study)

For each experimental session, a group of five participants were presented with and engaged in the same PRL task via an intranet connection without experimental deception. For a certain participant, portrait photos of the other four same-gender coplayers were always displayed within trials (Fig. 1A). This manipulation further increased the ecological validity of the task, at the same time creating a more engaging situation for the participants.

The social influence task contained six phases. (i) Phase 1. Initial choice (Choice 1). Upon the presentation of two choice options using abstract fractals, participants were asked to make their initial choice. A yellow frame was then presented to highlight the chosen option. (ii) Phase 2. Initial bet (Bet 1). After making Choice 1, participants were asked to indicate how confident they were in their choice, being "1" (not confident), "2" (reasonably confident), or "3" (very confident). Notably, the confidence ratings also served as postdecision wagering metric (an incentivized confidence rating) (*18*, *19*); namely, the ratings would be multiplied by their potential outcome on each trial. For instance, if a participant won on a particular trial, then the reward unit (i.e., 20 cents in the current setting) was multiplied with the rating (e.g., a bet of 2) to obtain the final outcome (20 × 2 = 40 cents). Therefore, the confidence rating in the current paradigm was referred to as bet. A yellow frame was presented to highlight the chosen bet. (iii) Phase 3. Preference giving. Once all participants had provided their Choice 1 and Bet 1, the choices (but not the bets) of the other coplayers were revealed. Crucially, instead of seeing all four other choices at the same time, participants had the opportunity to sequentially uncover their peer's decisions. In particular, participants could decide whom to uncover first and whom to uncover second, depending on their preference. Choices belonged to the preferred coplayers were then displayed underneath the corresponding photo. The remaining two choices were displayed automatically afterward. This manipulation was essential, because, in studies of decision-making, individuals tend to assign different credibility to their social peers based on their performance (*12*, *16*), and the resulting social preference may play an

important role in social decision-making (*23*). In the current study, because there were four other coplayers in the same learning environment, it was likely that they had various performance levels and therefore would receive different preferences from the observer. (iv) Phase 4. Choice adjustment (Choice 2). When all four other choices were presented, participants were able to adjust their choices given the instantaneous social information. The yellow frame was shifted accordingly to highlight the adjusted choice. (v) Phase 5. Bet adjustment (Bet 2). After the choice adjustment, participants might adjust their bet as well. In addition, participants also observed other coplayers' Choice 2 (on top of their Choice 1) once they had submitted their adjusted bets. Presenting other coplayers' choices after participants' bet adjustment rather than after their choice adjustment prevented a biased bet adjustment by the additional social information. The yellow frame was shifted accordingly to highlight the adjusted bet. (vi) Phase 6. Outcome delivery. Last, the outcome was determined by the combination of participants' Choice 2 and Bet 2 (e.g., $20 \times 2 = 40$ cents). Outcomes of the other coplayers were also displayed but shown only as of the single reward unit (i.e., 20 cents gain or loss) without being multiplied with their Bet 2. This was to provide participants with sufficient yet not overwhelming information about their peer's performance. On each trial, the reward was assigned to only one choice option given the reward probability; that is, only choosing one option would lead to a reward, whereas choosing the other option would lead to a punishment. The reward realization sequence (trial-by-trial complementary win and loss) was generated with a pseudo-random order according to the reward probability before the experiment, and this sequence was identical within each group.

Experimental procedure

To ensure a complete understanding of the task procedure, this study was composed of a 2-day procedure: prescanning training (day 1), and main experiment (day 2).

Prescanning training (day 1)

One to 2 days before the MRI scanning, five participants came to the behavioral laboratory to participate in the prescanning training. Upon arrival, they received the written task instruction and the consent form. After returning the written consent, participants were taken through a step-by-step task instruction by the experimenter. Notably, participants were explicitly informed (i) that an intranet connection was established so that they would observe real responses from the others, (ii) what probabilistic reward meant by receiving examples, (iii) that there was neither cooperation nor competition in this experiment, and (iv) that the reward probability could reverse multiple times over the course of the experiment, but participants were not informed about when and how often this reversal would take place. To shift the focus of the study away from social influence, we stressed the experiment as a multiplayer decision game, where the goal was to detect the "good option" so as to maximize their personal payoff in the end. Given this uncertainty, participants were instructed that they may either trust their own learning experience through trial and error or take decisions from their peers into consideration, as some of them might learn faster than the others. Participants' explicit awareness of all possible alternatives was crucial for the implementation of our social influence task. To further enhance participants' motivation, we informed them that the amount they would gain from the experiment would be added to their base payment (see the "Reward payment" section below). After participants had fully understood the task, we took portrait photos of them. To avoid emotional arousal, we asked participants to maintain a neutral facial expression as in typical passport photos. To prevent potential confusion before the training task, we further informed participants that they would only see photos of the other four coplayers without seeing themselves.

The training task contained 10 trials and differed from the main experiment in two aspects. First, it used a different set of stimuli than those used in the main experiment to avoid any learning effect. Second, participants were given a longer response window to fully understand every step of the task. Specifically, each trial began with the stimuli presentation of two choice alternatives, and participants were asked to decide on their Choice 1 (4000 ms) and Bet 1 (3000 ms). After the two sequential preference ratings (3000 ms each), all Choice 1 from the other four coplayers were displayed underneath their corresponding photos (3000 ms). Participants were then asked to adjust their choice (Choice 2; 4000 ms) and their bet (Bet 2; 3000 ms). Last, outcomes of all participants were released (3000 ms), followed by a jittered intertrial interval (ITI; 2000 to 4000 ms). To help participants familiarize themselves, we orally instructed them what to expect and what to do on each phase for the first two to three trials. The procedure during day 1 lasted about 1 hour.

Main experiment (day 2)

On the testing day, the five participants came to the MRI building. After a recap of all the important aspects of the task instruction, the MRI participant gave the MRI consent and entered the scanner to perform the main social influence task, while the remaining four participants were seated in the same room adjacent to the scanner to perform the task. All computers were interconnected via the intranet connection. They were further instructed not to make any verbal or gestural communications with other participants during the experiment.

The main experiment consisted of 100 trials and used a different pair of stimuli than the training task. It followed the exact description detailed above (see the "Breakdown of the social influence task (main study)" section and Fig. 1A). Specifically, each trial began with the stimulus presentation of two choice alternatives, and participants were asked to decide on their Choice 1 (2500 ms) and Bet 1 (2000 ms). After the two sequential preference ratings (2000 ms each), all Choice 1 from the other four coplayers were displayed underneath their corresponding photos (3000 ms). Participants were then asked to adjust their choice (Choice 2; 3000 ms) and their bet (Bet 2; 2000 ms). Last, outcomes of all participants were released (3000 ms), followed by a jittered ITI (2000 to 4000 ms). The procedure during day 2 lasted about 1.5 hours.

Reward payment

All participants were compensated with a base payment of 35 euros and the reward that they had achieved during the main experiment. In the main experiment, to prevent participants from careless responses on their Choice 1, they were explicitly instructed that on each trial, either their Choice 1 or their Choice 2 would be used to determine the final payoff. However, this did not affect the outcome delivery on the screen. Namely, although on some trials participants' Choice 1 was used to determine their payment, only outcomes that corresponded to their Choice 2 appeared on the screen. In addition, when their total outcome was negative, no money was deducted from their final payment. Overall, participants gained 4.48 ± 4.41 euros after completing the experiment. Last, the experiment ended with an informal debriefing session.

Behavioral data acquisition

Stimulus presentation, MRI pulse triggering, and response recording were accomplished with MATLAB R2014b (www.mathworks.com) and Cogent 2000 (www.vislab.ucl.ac.uk/cogent.php). In the behavioral group (as well as during the prescanning training), buttons "V" and "B" on the keyboard corresponded to the left and right choice options, respectively, and buttons V, B, and "N" corresponded to the bets 1, 2, and 3, respectively. As for the MRI group, a four-button MRI-compatible button box with a horizontal button arrangement was used to record behavioral responses. Buttons "a" and "b" on the button box corresponded to the left and right choice options, respectively, and buttons a, b, and "c" corresponded to the bets 1, 2, and 3, respectively. To avoid motor artifacts, the position of the two choices options was counterbalanced for all participants.

MRI data acquisition and preprocessing

MRI data collection was conducted on a Siemens Trio 3T scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. Each brain volume consisted of 42 axial slices (voxel size, $2 \times 2 \times 2$ mm^3, with 1-mm spacing between slices) acquired using a T2*-weighted echoplanar imaging (EPI) protocol (repetition time, TR = 2510 ms; echo time, TE = 25 ms; flip angle = 40°; field of view = 216 mm) in descending order. Orientation of the slice was tilted at 30° to the anterior commissure–posterior commissure (AC-PC) axis to improve signal quality in the orbitofrontal cortex (*42*). Data for each participant were collected in three runs with total volumes ranging from 1210 to 1230, and the first three volumes of each run were discarded to obtain a steady-state magnetization. In addition, a gradient echo field map was acquired before EPI scanning to measure the magnetic field inhomogeneity (TE1 = 5.00 ms, TE2 = 7.46 ms), and a high-resolution anatomical image (voxel size, $1 \times 1 \times 1$ mm^3) was acquired after the experiment using a T1-weighted MPRAGE protocol.

fMRI data preprocessing was performed using SPM12 (Statistical Parametric Mapping; Wellcome Trust Centre for Neuroimaging, University College London, London, UK). After converting raw Digital Imaging and Communications in Medicine (DICOM) images to NIfTI (Neuroimaging Informatics Technology Initiative) format, image preprocessing continued with slice timing correction using the middle slice of the volume as the reference. Next, a voxel displacement map (VDM) was calculated from the field map to account for the spatial distortion resulting from the magnetic field inhomogeneity (*43*, *44*). Incorporating this VDM, the EPI images were then corrected for motion and spatial distortions through realignment and unwarping. The participants' anatomical images were manually checked and corrected for the origin by resetting it to the AC-PC. The EPI images were then coregistered to this origin-corrected anatomical image. The anatomical image was skull-stripped and segmented into gray matter, white matter, and cerebrospinal fluid (CSF), using the "Segment" tool in SPM12. These gray and white matter images were used in the SPM12 DARTEL toolbox to create individual flow fields as well as a group anatomical template (*44*). The EPI images were then normalized to the MNI space using the respective flow fields through the DARTEL toolbox normalization tool. A Gaussian kernel of 6-mm full width at half maximum was used to smooth the EPI images.

After the preprocessing, we further identified brain volumes that (i) excessively deviated from the global mean of the blood oxygen level–dependent (BOLD) imaging signals (>1 SD), (ii) showed excessive head movement (movement parameter/TR > 0.4), or (iii) largely correlated with the movement parameters and the first derivative of the movement parameters ($R^2 > 0.95$). This procedure was implemented with the "Spike Analyzer" tool (https://github.com/GlascherLab/SpikeAnalyzer), which returned indices of those identified volumes. We then constructed them as additional participant-specific nuisance regressors of no interest across all our first-level analyses. This implementation identified 3.41 ± 4.79% of all volumes. Note that as this procedure was performed per participant, the total number of regressors for each participant may differ.

Behavioral data analysis

We tested for participants' behavioral adjustment after observing the instantaneous social information (during Phase 3), by assessing their choice switch probability in Phase 4 (how likely participants switched to the opposite option) and bet difference in Phase 5 (Bet 2 magnitude minus Bet 1 magnitude) as a measurement of how choice and confidence were modulated by the social information. Neither group difference (MRI versus behavioral) nor gender difference (male versus female) was observed for the choice switch probability (group: $F_{1,914} = 0.14$, $P = 0.71$; gender: $F_{1,914} = 0.24$, $P = 0.63$) and the bet difference (group: $F_{1,914} = 0.09$, $P = 0.76$; gender: $F_{1,914} = 1.20$, $P = 0.27$). Thus, we pulled data altogether to perform all subsequent analyses. In addition, trials where participants did not give valid responses on either Choice 1 or Bet 1 in time were excluded from the analyses. On average, 7.9 ± 7.3% of the entire trials were excluded.

We first tested how the choice switch probability and the bet difference varied as a function of the direction of the group (with and against, with respect to each participant's Choice 1) and the consensus of the group (2:2, 3:1, and 4:0, view of each participant; Fig. 1C) within trials. To this end, we submitted the choice switch probability and the bet difference to an unbalanced 2 (direction) × 3 (consensus) repeated-measures linear mixed-effect (LME) model. The unbalance was due to the fact that data in the 2:2 condition could only be used once, and we grouped it into the "against" condition, thus resulting in three consensus levels in the against condition and two consensus levels in the with condition. Grouping it into the with condition did not alter the results. Furthermore, we further tested the bet difference depending on whether participants switched or stayed on their Choice 2, by performing a 3 (group coherence, 2:2, 3:1, and 4:0) × 2 (direction, with versus against) × 2 (choice type, SwSt) repeated-measures LME models. We constructed LME models with different random effect specifications (table S1) and selected the best one for the subsequent statistical analyses (Fig. 1, D and E, and fig. S2A). We performed similar analyses with data from the nonsocial control study (fig. S1, C and D).

We further tested whether it was beneficial for the participants to adjust their choice and bet after receiving the instantaneous social information; that is, we assessed whether participants' switching behavior was elicited by considering social information or driven by purely perceptual mismatch (i.e., being confronted with visually distinct symbols). We reasoned that if participants were considering social information in our task, then the accuracy of their Choice 2 was expected to be higher than that of their Choice 1 (i.e., choosing the good option more often). By contrast, if participants' switching behavior was purely driven by perceptual mismatch, then a more random pattern ought to be expected, with no difference between the accuracy of Choice 1 and Choice 2. To this end, we assessed the difference in the accuracy between Choice 1 and Choice 2 (Fig. 1F), as well as the difference of the magnitude between Bet 1 and Bet 2

(Fig. 1G), using two-tailed paired *t* tests. We also tested how choice accuracy and bet magnitude changed across reversals. We selected a window of seven trials (three before and three after reversal, reversal included) to perform this analysis, with data being stacked with respect to the reversal (i.e., trial-locked) and averaged per participant. We submitted the data to a 2 (Choice 1 versus Choice 2 or Bet 1 versus Bet 2) × 7 (relative trial position, −3, −2, −1, 0, +1, +2, +3) repeated-measures LME models with five different random effect specifications, respectively (table S2). When the main effect of position was significant, we submitted the data to a post hoc comparison with Tukey's post hoc test correction (fig. S2, B and C). We performed similar analyses with data from the nonsocial control study (fig. S1, E and F).

In addition, although we did not intentionally manipulate the amount of dissenting social information in the main experiment (given the real-time property of our task), the sequence was nonetheless randomly distributed for nearly all participants (Wald-Wolfowitz runs test, all $P > 0.05$). To guard against possible confounding effects, we nonetheless tested whether the amount of dissenting social information and participants' behavior was related to task structure (time of reversal) and participants' lapse error. Note that the lapse error was defined as choosing one choice option on Choice 1 when the model strongly favored the alternative (modeled action probability ≥ 95%). For example, when the model predicted $P(\mathrm{A})$ of Choice 1 was 95% (or higher) yet the participants actually chose option B, this trial was referred to as a lapse error. We tested the Pearson's correlation between the following pairs of variables for all participants and for MRI participants: (i) amount of dissenting social information and time of reversal, (ii) amount of dissenting social information and lapse error, (iii) participants' switching behavior and time of reversal, and (iv) participants' switching behavior and lapse error. Results indicated no significant relationship between any of the above pairs of variables (fig. S2, E and F).

Last, we tested how choice accuracy and bet magnitude changed between trials, as a function of choice adjustment on the previous trial (Choice 2 SwSt), the relative direction of the group (with versus against), and the group consensus (2:2, 3:1, 4:0). That is, we assessed the carry-over effect after participants had observed the others' Choice 2 behavior. To this end, we submitted the choice accuracy and the bet magnitude to an unbalanced 2 (adjustment) × 2 (direction) × 3 (consensus) repeated-measures LME model. The unbalance was due to the fact that data in the 2:2 condition could only be used once, and we grouped it into the against condition. Grouping it into the with condition did not alter the results. We constructed LME models with different random effect specifications (table S1) and selected the best one for the subsequent statistical analyses (Fig. 1, H and I). We performed similar analyses with data from the nonsocial control study (fig. S1, H and I).

All statistical tests were performed in R (v3.3.1; www.r-project.org). All repeated-measures LME models were analyzed with the "lme4" package in R. Results were considered statistically significant at the level $P < 0.05$.

Computational modeling

To describe participants' learning behavior in our social influence task and to uncover latent trial-by-trial measures of decision variables, we developed three categories of computational models and fitted these models to participants' behavioral data. We based all our computational models on the simple RL model (*5*) and progressively include components (Table 1).

First, given the structure of the PRL task, we sought to evaluate whether a fictitious update RL model (*20*) that incorporates the anticorrelation structure (see the "Underlying PRL paradigm" section) outperformed the simple Rescorla-Wagner (*21*) RL model that only updated the value of the chosen option and the Pearce-Hall (*45*) model that used a dynamic learning rate to approximate the optimal Bayesian learner. These models served as the baseline and did not consider any social information (category 1: M1a, M1b, and M1c). On top of category 1 models, we then included the instantaneous social influence (i.e., other coplayers' Choice 1, before outcomes were delivered) to construct social models (category 2: M2a, M2b, and M2c). Last, we considered the component of social learning with competing hypotheses of value update from observing others (category 3: M3, M4, M5, M6a, and M6b). The remainder of this section explains choice-related model specifications and bet-related model specifications (see table S3 for a list of full specifications). All models were estimated and evaluated under the hierarchical Bayesian framework (note S2).

Choice model specifications

In all models, Choice 1 was accounted for by the option values of option A and option B

$$\mathbb{V}_t = [V_t(\mathrm{A}), V_t(\mathrm{B})] \tag{3}$$

where V_t indicated a two-element vector consisting of option values of A and B on trial *t*. Values were then converted into action probabilities using a Softmax function (*5*). On trial *t*, the action probability of choosing option A (between A and B) was defined as follows

$$P_t(\mathrm{A}) = \frac{e^{V_t(\mathrm{A})}}{e^{V_t(\mathrm{A})} + e^{V_t(\mathrm{B})}} = \frac{1}{1 + e^{-(V_t(\mathrm{A}) - V_t(\mathrm{B}))}} \tag{4}$$

For Choice 2, we modeled it as a "switch" (coded as 1) or a "stay" (coded as 0) using a logistic regression. On trial *t*, the probability of switching given the switch value was defined as follows

$$P_t(\mathrm{switch}) = \Phi(V_t(\mathrm{switch})) \tag{5}$$

where Φ was the inverse logit linking function

$$\Phi(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

Note that, in model specifications of the action probability, we did not include the commonly used inverse Softmax temperature parameter τ. This was because we explicitly constructed the option values of Choice 1 and the switch value of Choice 2 in a design-matrix fashion (e.g., Eq. 8; see the text below). Therefore, including the inverse Softmax temperature parameter would inevitably give rise to a multiplication term, which, as a consequence, would cause unidentifiable parameter estimation (*24*). For completeness, we also assessed models with the τ parameter, and they performed consistently worse than our models specified here.

The category 1 models (M1a, M1b, and M1c) did not consider any social information. In the simplest model (M1a), a Rescorla-Wagner model was used to model the Choice 1, with only the chosen value being updated via the RPE (δ), and the unchosen value remaining the same as the last trial.

$$\begin{aligned} \delta_{\mathrm{chosen},C2,t} &= R_t - V_{\mathrm{chosen},C2,t} \\ V_{\mathrm{chosen},C2,t+1} &= V_{\mathrm{chosen},C2,t} + \alpha\delta_{\mathrm{chosen},C2,t} \\ V_{\mathrm{unchosen},C2,t+1} &= V_{\mathrm{unchosen},C2,t} \end{aligned} \tag{7}$$

where R_t was the outcome on trial t and α ($0 < \alpha < 1$) denoted the learning rate that accounted for the weight of RPE in value update. A β weight (β_V) was then multiplied with the values before being submitted to Eq. 4 with a categorical distribution, as in

$$C1_t \sim \text{Categorical}(\text{Softmax}(\beta_v \mathbb{V}_t)) \quad (8)$$

Because there was no social information in M1a, the switch value of Choice 2 was composed merely of the value difference of Choice 1 and a switching bias (i.e., intercept)

$$V_t(\text{switch}) = \beta_{\text{bias}_{C2}} + \beta_{\text{vdiff}_{C2}}(V_{\text{chosen},C1,t} - V_{\text{unchosen},C1,t}) \quad (9)$$

Choice 2 was then modeled with this switch value following a Bernoulli distribution

$$C2 \sim \text{Bernoulli}(V_t(\text{switch})) \quad (10)$$

In M1b, we tested whether the fictitious update could improve the model performance, as the fictitious update has been successful in PRL tasks in nonsocial contexts (*20*, *27*). In M1b, both the chosen value and the unchosen value were updated, as in

$$\begin{aligned} \delta_{\text{chosen},C2,t} &= R_t - V_{\text{chosen},C2,t} \\ \delta_{\text{unchosen},C2,t} &= -R_t - V_{\text{unchosen},C2,t} \\ V_{\text{chosen},C2,t+1} &= V_{\text{chosen},C2,t} + \alpha\delta_{\text{chosen},C2,t} \\ V_{\text{unchosen},C2,t+1} &= V_{\text{unchosen},C2,t} + \alpha\delta_{\text{unchosen},C2,t} \end{aligned} \quad (11)$$

In M1c, we assessed the Pearce-Hall (*45*) model that entailed a dynamic learning rate

$$\begin{aligned} \delta_{\text{chosen},C2,t} &= R_t - V_{\text{chosen},C2,t} \\ V_{\text{chosen},t+1} &= V_{\text{chosen},t} + k\alpha_t\delta_{\text{chosen},C2,t} \\ V_{\text{unchosen},C2,t+1} &= V_{\text{unchosen},C2,t} \\ \alpha_{t+1} &= \lambda \mid \delta_{\text{chosen},t} \mid + (1-\lambda)\alpha_t \end{aligned} \quad (12)$$

where k ($0 < k < 1$) was the weight of the (dynamic) learning rate and λ ($0 < \lambda < 1$) indicated the weight between RPE and the learning rate.

Our category 2 models (M2a, M2b, and M2c) tested the role of instantaneous social influence on Choice 2, namely, whether observing choices from the other coplayers contributed to the choice switching. As compared with M1 (M1a, M1b, and M1c), only the switch value of Choice 2 was modified, as follows

$$\begin{aligned} V_t(\text{switch}) = \beta_{\text{bias}_{C2}} &+ \beta_{\text{vdiff}_{C2}}(V_{\text{chosen},C1,t} - V_{\text{unchosen},C1,t}) \\ &+ \beta_{\text{against}}\, w.\text{N}_{\text{against},t} \end{aligned} \quad (13)$$

where $w.\text{N}_{\text{against},t}$ denoted the preference-weighted amount of dissenting social information relative to each participant's Choice 1 on trial t. It was computed on a trial-by-trial fashion as follows

$$w.\text{N}_{\text{against},t} = \frac{\sum_{s=1}^{K} w_{s,t}}{\sum_{s=1}^{4} w_{s,t}}, K = 0,1,\ldots,4 \quad (14)$$

where K indicated the number of opposite choices from the others and $w_{s,t}$ was participants' trial-by-trial preference weight toward the other four coplayers. Note that these preference weights were fixed parameters based on each participant's preference toward the others when uncovering their choices: The first favored coplayer received a weight of 0.75, the second favored coplayer received a weight of 0.5, and the remaining two coplayers received a weight of 0.25, respectively. They were not modeled as free parameters because doing so caused unidentifiable model estimate behavior. All other specifications of models in this category (M2a, M2b, and M2c) were identical to models in category 1 (M1a, M1b, and M1c), respectively.

Our category 3 models (M3, M4, M5, M6a, and M6b) assessed whether participants learned from their social partners and whether they updated vicarious option values through social learning. Note that models belonging to category 2 solely considered the instantaneous social influence on Choice 2, whereas models in category 3 tested several competing hypotheses of the vicarious valuation that may contribute to Choice 1 on the following trial, in combination with individuals' own valuation processes. In all models within this category, the option values of Choice 1 were specified by a weighted combination between V_{self} updated via direct learning and V_{other} updated via social learning

$$\mathbb{V}_t = \beta_{\text{vself}}\mathbb{V}_{\text{self},t} + \beta_{\text{vother}}\mathbb{V}_{\text{other},t} \quad (15)$$

where

$$\begin{aligned} \mathbb{V}_{\text{self},t} &= [V_{\text{self},t}(\text{A}), V_{\text{self},t}(\text{B})] \\ \mathbb{V}_{\text{other},t} &= [V_{\text{other},t}(\text{A}), V_{\text{other},t}(\text{B})] \end{aligned} \quad (16)$$

Note that given that M2b was the winning model among category 1 and category 2 models (Table 1), we used M2b's specification for the value update of V_{self} (Eq. 11), so that category 3 models only differed on the specification of V_{other}.

M3 tested whether individuals recruited a similar RL algorithm to their own when learning option values from observing others. Hence, M3 assumed participants to update values "for" the others using the same fictitious update rule for themselves

$$\begin{aligned} \delta_{s,\text{chosen},C2,t} &= R_{s,t} - V_{s,\text{chosen},C2,t}, s = 1,2,3,4 \\ \delta_{s,\text{unchosen},C2,t} &= -R_{s,t} - V_{s,\text{unchosen},C2,t} \\ V_{s,\text{chosen},C2,t+1} &= V_{s,\text{chosen},C2,t} + \alpha_o\delta_{s,\text{chosen},C2,t} \\ V_{s,\text{unchosen},C2,t+1} &= V_{s,\text{unchosen},C2,t} + \alpha_o\delta_{s,\text{unchosen},C2,t} \end{aligned} \quad (17)$$

where s denoted the index of the four other coplayers and α_o was the learning rate for the others. These option values from the four coplayers were then preference-weighted and summed to formulate V_{other}, as follows

$$\begin{aligned} V_{\text{other},t+1}(\text{A}) &= \sum_{s=1}^{4} w_{s,t}\, V_{s,t+1}(\text{A}) \\ V_{\text{other},t+1}(\text{B}) &= \sum_{s=1}^{4} w_{s,t}\, V_{s,t+1}(\text{B}) \end{aligned} \quad (18)$$

where $w_{s,t}$ was participants' preference weight. To ensure that the corresponding value-related parameters (β_{vself} and β_{vother} in Eq. 15) were comparable, V_{other} was further normalized to lie between −1 and 1 with the $\Phi(x)$ function defined in Eq. 6

$$\begin{aligned} V_{\text{other},t+1}(\text{A}) &= 2\Phi(V_{\text{other},t+1}(\text{A})) - 1 \\ V_{\text{other},t+1}(\text{B}) &= 2\Phi(V_{\text{other},t+1}(\text{B})) - 1 \end{aligned} \quad (19)$$

One may argue that having four independent RL agents as in M3 was cognitively demanding: To accomplish so, participants had to

track and update each other's individual learning processes together with their own valuation (together 2^5 units of information). We, therefore, constructed three additional models that used simpler but distinct pathways to update vicarious values via social learning. In essence, M3 considered both choice and outcome to determine the action value. We then asked whether using either choice or outcome alone may perform well as, or even better than, M3. Following this assumption, we constructed (i) M4 that updated V_{other} using only the others' action preference, (ii) M5 that considered the others' current outcome to resemble the value update via observational learning, and (iii) M6a that tracked the others' cumulative outcome to resemble the value update via observational learning.

In M4, other players' action preference (ρ) is derived from the choice history over the last three trials using the cumulative distribution function of the beta distribution at the value of 0.5 ($I_{0.5}$). That is

$$\begin{aligned}\rho_{s,t}(\mathrm{A}) &= I_{0.5}(1+\sum_{t=T-2}^{T} C2_{\mathrm{B},s,t}, 1+\sum_{t=T-2}^{T} C2_{\mathrm{A},s,t})\\ \rho_{s,t}(\mathrm{B}) &= 1-\rho_{s,t}(\mathrm{A})\end{aligned} \tag{20}$$

where s denoted the index of the four other coplayers and t denoted the trial index from $T-2$ to T. To illustrate, if one coplayer chose option A twice and option B once in the last three trials, then the action preference of choosing A for him/her was as follows: $I_{0.5}$(frequency of B + 1, frequency of A + 1) = $I_{0.5}$(0.5, 1 + 1, 2 + 1) = 0.6875. V_{other} was computed on the basis of these action preferences

$$\begin{aligned}V_{\mathrm{other},t+1}(\mathrm{A}) &= \sum_{s=1}^{4} w_{s,t}\rho_{s,t}(\mathrm{A})\\ V_{\mathrm{other},t+1}(\mathrm{B}) &= \sum_{s=1}^{4} w_{s,t}\rho_{s,t}(\mathrm{B})\end{aligned} \tag{21}$$

where $w_{s,t}$ was participants' preference weight and s denoted the index of the four other coplayers. Similar to M3, the computation of V_{other} here was also preference-weighted and summed. The values were similarly normalized using Eq. 19.

By contrast, M5 tested whether participants updated V_{other} using only each other's reward on the current trial, which was equivalent to the standard Rescorla-Wagner model with $\alpha = 1$, indicating no trial-by-trial learning

$$\begin{aligned}V_{\mathrm{other},t+1}(\mathrm{A}) &= \sum_{s=1}^{K_\mathrm{A}} w_{s,t}R_{s,t}, K_\mathrm{A} = 0,1,\ldots,4\\ V_{\mathrm{other},t+1}(B) &= \sum_{s=1}^{4-K_\mathrm{A}} w_{s,t}R_{s,t}\end{aligned} \tag{22}$$

where $w_{s,t}$ was participants' preference weight, s denoted the index of the four other coplayers, t denoted the trial index from $T-2$ to T, and K_A denoted the number of coplayers who decided on option A on trial t. Similar to M3, the computation of V_{other} here was also preference-weighted and summed. These values were then normalized using Eq. 19.

Moreover, M6a assessed whether participants tracked cumulated reward histories over the last few trials instead of monitoring only the most recent outcome of the others. A discounted reward history over the recent past (e.g., the last three trials) has been a relatively common implementation in other RL studies in nonsocial contexts (*22*, *46*). By testing four window sizes of trials (i.e., two, three, four, or five) and using a nested model comparison, we decided on a window of three past trials to accumulate the other coplayers' performance

$$\begin{aligned}V_{\mathrm{other},t+1}(\mathrm{A}) &= \sum_{s=1}^{K_\mathrm{A}}\sum_{t=T-2}^{T} w_{s,t}\gamma^{T-i}R_{s,i}, K_\mathrm{A} = 0,1,\ldots,4\\ V_{\mathrm{other},t+1}(\mathrm{B}) &= \sum_{s=1}^{4-K_\mathrm{A}}\sum_{t=T-2}^{T} w_{s,t}\gamma^{T-i}R_{s,i}\end{aligned} \tag{23}$$

where γ ($0 < \gamma < 1$) denoted the rate of exponential decay and all other notions were as in Eq. 22. Similar to M3, the computation of V_{other} here was also preference-weighted and summed. The values were then normalized using Eq. 19.

Last, given that M6a was the winning model among all the models above (M1 to M6a) indicated by model comparison (see below model selection; Table 1), we further assessed in M6b whether Bet 1 contributed to the choice switching on Choice 2, as follows

$$\begin{aligned}V_t(\mathrm{switch}) = \beta_{\mathrm{bias}_{C2}} &+\beta_{\mathrm{vdiff}_{C2}}(V_{\mathrm{chosen},C1,t} - V_{\mathrm{unchosen},C1,t})\\ &+\beta_{\mathrm{against}}w.\mathrm{N}_{\mathrm{against},t} + \beta_{\mathrm{bet1}}\mathrm{Bet}\,1_t\end{aligned} \tag{24}$$

Note that in M6a/M6b, V_{other} differed from V_{self} in practice. On trial t, V_{self} of a punished option might largely decrease given the negative RPE, whereas V_{other} may not be vastly affected because of the others' previous successes [e.g., V_{other}(blue); Fig. 2C; albeit a loss on trial t, the cumulative reward history was still positive, indicating that the cumulative performance was still reliable]. Both V_{self} and V_{other} spanned within their range (−1 to 1; Fig. 2D) with a slightly moderate correlation ($r = 0.38 \pm 0.097$ across participants; Fig. 3A), and they jointly contributed to the action probability of Choice 1.

Bet model specifications

In all models, both Bet 1 and Bet 2 were modeled as ordered-logistic regressions that are often used for quantifying ordered discrete variables, such as Likert-scale questionnaire data (*24*). We applied the ordered-logistic model because the bets in our study indeed inferred an ordinal feature. Namely, betting on three was higher than betting on two, and betting on two was higher than betting on one, but the difference between the bets of three and one (i.e., a difference of two) was not necessarily twice as the difference between the bets of three and two (i.e., a difference of one). Hence, we sought to model the distance (decision boundary) between them. Moreover, we hypothesized a continuous computation process of bet utilities when individuals were placing bets, which satisfied the general assumption of the ordered-logistic regression model.

There were two key components in our bet models, the continuous bet utility U_{bet} and the set of boundary thresholds θ. Specifically, the bet utility U_{bet} varied between $K-1$ thresholds ($\theta_{1,2,\ldots,K-1}$) thresholds to predict bets. Since there were three bet levels in our task ($K = 3$), we introduced two decision thresholds, θ_1 and θ_2 (where $\theta_2 > \theta_1$). Hence, the predicted bets (bêt) on trial t were represented as follows

$$\hat{\mathrm{bet}}_{i,t} = \begin{cases}1, \text{if } -\infty < U_{\mathrm{bet}_{i,t}} < \theta_1\\ 2, \text{if } \theta_1 < U_{\mathrm{bet}_{i,t}} < \theta_2, i = 1,2\\ 3, \text{if } \theta_2 < U_{\mathrm{bet}_{i,t}} < +\infty\end{cases} \tag{25}$$

where i indicated either Bet 1 or Bet 2. Because there were only two levels of threshold, for simplicity, we set $\theta_1 = 0$ and $\theta_2 = \theta$

(where $\theta > 0$). To model the actual bets, a logistic function (Eq. 6) was used to obtain the action probability of each bet, as follows

$$\begin{cases} P(\text{bet}_{i,t} = 1) = \Phi(-U_{\text{bet}_{i,t}}) \\ P(\text{bet}_{i,t} = 2) = \Phi(\theta - U_{\text{bet}_{i,t}}) - \Phi(-U_{\text{bet}_{i,t}}), i = 1,2 \\ P(\text{bet}_{i,t} = 3) = 1 - \Phi(\theta - U_{\text{bet}_{i,t}}) \end{cases} \quad (26)$$

The utility U_{bet1} was composed of a bet bias and the value difference between the chosen option and the unchosen option

$$U_{\text{bet1},t} = \beta_{\text{bias}_{B1}} + \beta_{\text{vdiff}_{B1}}(V_{\text{chosen},C1,t} - V_{\text{unchosen},C1,t}) \quad (27)$$

The rationale was that the larger the value difference between the chosen and the unchosen options, the more confident individuals were expected to be, hence placing a higher bet. This utility U_{bet1} was kept identical across all models (M1a to M6b), and Bet 1 was modeled as follows

$$B1_t \sim \text{OrderedLogistic}(U_{\text{bet1},t} \mid \theta) \quad (28)$$

In addition, Bet 2 was modeled as the bet change relative to Bet 1. Therefore, the utility U_{bet2} was constructed on the basis of U_{bet1}. In all nonsocial models (M1a, M1b, and M1c), the bet change term was represented by a bet change bias (i.e., intercept), depending on whether participants had a switch or stay on their Choice 2

$$U_{\text{bet2},t} = \begin{cases} U_{\text{bet1},t} + \beta_{\text{bias}_{\text{stay}}}, \text{if } C1 = C2 \\ U_{\text{bet1},t} + \beta_{\text{bias}_{\text{switch}}}, \text{if } C1 \neq C2 \end{cases} \quad (29)$$

In all social models (M2a to M6b), regardless of the observational learning effect, the bet change term was specified by the instantaneous social information together with the bias, depending on whether participants had a switch or stay on their Choice 2

$$U_{\text{bet2},t} = \begin{cases} U_{\text{bet1},t} + \beta_{\text{with}_{\text{stay}}} w.\text{N}_{\text{with},t} + \beta_{\text{against}_{\text{stay}}} w.\text{N}_{\text{against},t}, \text{if } C1 = C2 \\ U_{\text{bet1},t} + \beta_{\text{with}_{\text{switch}}} w.\text{N}_{\text{with},t} + \beta_{\text{against}_{\text{switch}}} w.\text{N}_{\text{against},t}, \text{if } C1 \neq C2 \end{cases} \quad (30)$$

with

$$w.\text{N}_{\text{against},t} = \frac{\sum_{s=1}^{K} w_{s,t}}{\sum_{s=1}^{4} w_{s,t}}, K = 0,1,\ldots,4$$
$$w.\text{N}_{\text{with},t} = \frac{\sum_{s=1}^{4-K} w_{s,t}}{\sum_{s=1}^{4} w_{s,t}} \quad (31)$$

where K indicated the number of opposite choices from the others and $w_{s,t}$ was participants' trial-by-trial preference weight toward the other four coplayers. Note that, however, despite the high negative correlation between $w.\text{N}_{\text{with}}$ and $w.\text{N}_{\text{against}}$, the parameter estimation results showed that the corresponding effects (i.e., β_{with} and β_{against}) did not rely on each other ($r = 0.04$, $P > 0.05$). As shown in Fig. 2H, the corresponding parameters showed independent contributions to bet changes during the adjustment. In addition, we constructed two other models using either $w.\text{N}_{\text{with}}$ or $w.\text{N}_{\text{against}}$ alone, but both models' performance markedly reduced than including both of them [ΔLOOIC (leave-one-out information criterion relative to the winning model) > 1000]. Last, the utility U_{bet2} was kept identical across all social models (M2a to M6b), and Bet 2 was modeled as follows

$$B2_t \sim \text{OrderedLogistic}(U_{\text{bet2},t} \mid \theta) \quad (32)$$

MRI data analysis

Deriving internal computational signals

On the basis of the winning model (Table 1) and its parameter estimation (Fig. 2, E to H), we derived trial-by-trial computational signals for each MRI participant using the mean of the posterior distribution of the parameters. We used the mean rather than the mode (i.e., the peak resulted from kernel density estimate) because in Markov chain Monte Carlo, especially Hamiltonian Monte Carlo (HMC) implemented in Stan, the mean is much more stable than the mode to serve as the point estimate of the entire posterior distribution (*24*). As we modeled all parameters as normal distributions, the posterior mean and the posterior mode were highly correlated ($r = 0.99$, $P < 1.0 \times 10^{-10}$). For each MRI participant, we derived the following trial-by-trial variables and behaviors: V_{self}, V_{other}, $w.\text{N}_{\text{against}}$, Choice 2 behavior (SwSt), U_{bet1}, U_{bet2}, and RPE.

First-level analysis

fMRI data analyses were performed using SPM12. We conducted model-based fMRI analyses (*20*, *25*) containing the computational signals described above. We set up two event-related GLMs (GLM 1 and GLM 2) to test the neural correlates of decision variables.

GLM 1 assessed the neural representations of valuation resulted from participants' direct learning and observational learning in Phase 1, as well as the instantaneous social influence in Phase 3. The first-level design matrix in GLM 1 consisted of constant terms, nuisance regressors identified by the Spike Analyzer, and the following 22 regressors: 5 experimentally measured onset regressors for each cue (cue of Choice 1: 0 s after trial began; cue of Bet 1: 2.92 s after trial began; cue of Choice 2: 12.82 s after trial began; cue of Bet 2: 16.25 s after trial began; cue of outcome: 21.71 s after trial began; all the timing here corresponded to the mean onsets for each cue across trials and participants); 6 parametric modulators (PMs) of each corresponding cue ($V_{\text{self,chosen}}$ and $V_{\text{other,chosen}}$, belonging to the cue of Choice 1; $w.\text{N}_{\text{against}}$ belonging to the cue of Choice 2; U_{bet1} and U_{bet2}, belonging to the cue of Bet 1 and Bet 2, respectively; and RPE belonging to the cue of outcome); 5 nuisance regressors accounted for all of the "no-response" trials (missing trials) of each cue; and 6 movement parameters. Note that $V_{\text{other,chosen}}$ was orthogonalized with respect to $V_{\text{self,chosen}}$. This allowed us to obtain as much variance as possible on the $V_{\text{self,chosen}}$ regressor, and then any additional (explainable) variance would be accounted for by the $V_{\text{other,chosen}}$ regressor (*47*). In addition, we intentionally did not include the actual reward outcome at the outcome cue. This was because (i) the RPE and the reward outcome are known to be correlated in goal-directed learning studies using model-based fMRI and (ii) we sought to explicitly verify RPE signals by its hallmarks using the region of interest (ROI) time series extracted from each participant

given the second-level RPE contrast (see the "Follow-up ROI analysis" section).

GLM 2 was set up to examine the neural correlates of choice adjustment in Phase 4. To this end, GLM 2 was identical to GLM 1, except that the PM regressor of $w.N_{against}$ under the onset cue of Choice 2 was replaced by the PM regressor of SwSt (switch = 1, stay = −1). In addition, albeit that we showed no pattern between participants' behavior and task structure (fig. S2, E and F), we included each participant's time of reversal and their lapse error as covariates in GLM 1 and GLM 2, resulting in two new GLMs, GLM 3 and GLM 4. Given the noncorrelation between variables of interest and the task structure, significant clusters resulted from GLM 3 and GLM 4 were nearly identical with those from GLM 1 and GLM 2, respectively.

Second-level analysis

The resulting β images from each participant's first-level GLMs were then used in random-effects group analyses at the second level, using one-sample two-tailed *t* tests for significant effects across participants. To correct for multiple comparisons of functional imaging data, we used the threshold-free cluster enhancement (TFCE) (*48*) implemented in the TFCE Toolbox (dbm.neuro.uni-jena.de/tfce/). TFCE is a cluster-based thresholding method that aims to overcome the shortcomings of choosing an arbitrary cluster size (e.g., $P < 0.001$, cluster size $k = 20$) to form a threshold. The TFCE procedure took the raw statistics from the second-level analyses and performed a permutation-based nonparametric test (i.e., 5000 permutations in the current study) to obtain robust results. In addition, given our hypotheses and according to existing evidence that vmPFC encodes experiential value signals from direct learning (*9*) and that ACC tracks vicarious value signals from social learning (*8*, *12*, *32*), we performed small volume corrections for the value related contrast using 10-mm search volumes around the peak MNI coordinates of the vmPFC ($x = 2$, $y = 46$, $z = −8$) and the ACC ($x = 2$, $y = 14$, $z = 30$) reported in the corresponding studies with the TFCE correction at $P < 0.05$ (Fig. 3B). For the otherwise whole-brain analyses, we performed whole-brain TFCE correction at $P < 0.05$, FWE (family-wise error) corrected (Fig. 3C and figs. S4 and S5).

Follow-up ROI analysis

Depending on the hypotheses, the research question, and the corresponding PM regressors, we used two types of follow-up ROI analyses: the time series estimates and percent signal change (PSC) estimates. In both types of ROI analyses, participant-specific masks were created from the second-level contrast. We applied a previously reported leave-one-out procedure (*20*) to extract cross-validated BOLD time series. This was to provide an independent criterion for ROI identification and thus ensured statistical validity. For each participant, we first defined a 10-mm search volume around the peak coordinate of the second level contrast re-estimated from the remaining $n − 1$ participants (threshold: $P < 0.001$, uncorrected); within this search volume, we then searched for each participant's nearest individual peak and created a new 10-mm sphere around this individual peak as the ROI mask. Last, suprathreshold voxels in the new participant-specific ROI were used for both ROI analyses.

The ROI time series estimates were applied when at least two PMs were associated with each ROI. Namely, we were particularly interested in how the time series within a specific ROI correlated with all the PM regressors. In the current studies, we defined three ROIs to perform the time series estimates: vmPFC, ACC, and VS/NAcc.

We followed the procedure established by previous studies (*12*, *26*) to perform the ROI time series estimates. We first extracted raw BOLD time series from the ROIs. The time series of each participant was then time-locked to the beginning of each trial with a duration of 30 s, where the cue of Choice 1 was presented at 0 s, the cue of Bet 1 was presented at 2.92 s, the cue of Choice 2 was displayed at 12.82 s, the cue of Bet 2 was displayed at 16.25 s, and the cue of outcome was presented at 21.71 s. Afterward, ROI time series were upsampled to a resolution of 250 ms ($^1/_{10}$ of TR) using two-dimensional cubic spline interpolation, resulting in a data matrix of size $m \times n$, where m was the number of trials and n was the number of the upsampled time points (i.e., 30 s/250 ms = 120 time points). A linear regression model containing the PMs was then estimated at each time point (across trials) for each participant. Note that, although the linear regression here took a similar formulation as the first-level GLM, it did not model any specific onset; instead, this regression was fitted at each time point within the entire trial across all trials. The resulting time courses of effect sizes (regression coefficients or β weights) were lastly averaged across participants.

To test the group-level significance of the above ROI time series analysis, we used a nonparametric permutation procedure. For the time sources of effect sizes (β weights) for each ROI, we defined a time window of 3 to 7 s after the corresponding event onset, during which the BOLD response was expected to peak. In this time window, we randomly flipped the signs of the time courses of β weights for 5000 repetitions to generate a null distribution and assessed whether the mean of the generated data from the permutation procedure was smaller or larger than 97.5% of the mean of the empirical data, as the *P* value.

Further, the PSC estimates were applied when only one PM was associated with each ROI. In particular, we tested whether there was a linear trend of the PSC for each ROI as a function of the PM. In the current study, we defined seven ROIs to perform the PSC estimates. Among them, four ROIs were associated with the PM regressor of $w.N_{against}$, being the rTPJ, the ACC/posterior medial frontal cortex (pMFC), the right anterior insula (aINS), and the frontopolar cortex (FPC); two ROIs were associated with the PM regressor of SwSt, being the left dlPFC and the ACC/pMFC; and one ROI was associated with the inverse contrast of SwSt (i.e., stay versus switch), being the vmPFC.

To compute the PSC, we used the "rfxplot" toolbox (*49*) to extract the time series from the above ROIs. The rfxplot toolbox further divided the corresponding PMs into different bins (e.g., in the case of two bins, PMs were split into the first 50% and the second 50%) and computed the PSC for each bin, which resulted in a $p \times q$ PSC matrix, where p was the number of participants and q was the number of bins. To test for significance, we performed a simple first-order polynomial fit using the PSC as a function of the binned PM and tested whether the slope of this polynomial fit was significantly different from zero using two-tailed one sample *t* tests.

Functional connectivity analysis

We conducted two types of functional connectivity analyses (*28*) in the current study, the PPI and the PhiPI, to assess the functional network using fMRI. In both types of connectivity analyses, the seed brain regions were determined on the basis of the activations from the earlier GLM analyses, and we extracted cross-validated BOLD time series from each corresponding ROI using the leave-one-out procedure described above.

The PPI analysis aims to uncover how the functional connectivity between BOLD signals in a particular ROI (seed region) and

BOLD signals in the (to-be-detected) target region(s) is modulated by a psychological variable. We used as a seed the entire BOLD time series from a 10-mm spherical ROI in the rTPJ, centered at the peak coordinates from the PM contrast of $w.N_{against}$ (threshold: $P < 0.001$, uncorrected), which was detected at the onset cue of the second choice. Next, we constructed the interaction regressor of the PPI analysis (i.e., the regressor of main interest) by combining the rTPJ ROI signals with the SwSt (switch = 1, stay = −1) variable that took place at the onset cue of Choice 2. We first normalized the physiological and psychological terms and then multiplied them together, further orthogonalizing their product to each of the two main effects. These three regressors (i.e., the interaction, the BOLD time series of the seed region, and the modulating psychological variable) were lastly mean-corrected and then entered into the first-level PPI design matrix. To avoid possible confounding effects, we further included all the same nuisance regressors as the above first-level GLMs: five nuisance regressors accounted for all the no-response trials (missing trials) of each event cue, six movement parameters, and additional regressors of interest identified by the Spike Analyzer. The resulting first-level interaction regressor from each participant was then submitted to a second-level *t* test to establish the group-level connectivity results, with whole-brain TFCE correction at $P < 0.05$, FWE corrected (Fig. 4, A to C).

The PhiPI analysis follows the same principles as the PPI analysis, except that the psychological variable in the PPI regressors is replaced by the BOLD time series from a second seed ROI. For the interaction term, we first normalized the BOLD time series of the two seed regions and then multiplied them together, further orthogonalizing their product to each of the two main effects. The three regressors (i.e., two main-effect terms and their interaction) were lastly mean-corrected and then entered into the first-level PhiPI design matrix.

We performed two PhiPI analyses. In the first PhiPI, we used as seed regions the entire BOLD time series in two 10-mm spherical ROIs in the vmPFC (seed 1) and the ACC (seed 2), both of which were detected at the cue of Choice 1 from the PMs of V_{self} and V_{other}, respectively. The design matrix of the first PhiPI analysis thus consisted of the interaction term between vmPFC and ACC and the two main-effect regressors with the BOLD time series of vmPFC and ACC, respectively. In the second PhiPI, we seeded with the entire BOLD time series from an identical 10-mm spherical ROI in the rTPJ (seed 1) as described in the above PPI analysis and from a 10-mm spherical ROI in the left dlPFC (seed 2), which was identified at the cue of Choice 2 from the contrast of choice adjustment (switch > stay). The design matrix of the second PhiPI analysis thus consisted of the interaction term between rTPJ and left dlPFC and the two main-effect regressors with the BOLD time series of rTPJ and left dlPFC, respectively. In both PhiPI analyses, we further included all the same nuisance regressors as the above first-level GLMs to avoid possible confounding effects. The resulting first-level interaction regressor from each participant was then submitted to a second-level *t* test to establish the group-level connectivity results, with whole-brain TFCE correction at $P < 0.05$, FWE corrected (Fig. 4, E to I, and fig. S6A).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/34/eabb4159/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. S. E. Asch, Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr. Gen. Appl.* **70**, 1–70 (1956).
2. V. Klucharev, K. Hytönen, M. Rijpkema, A. Smidts, G. Fernández, Reinforcement learning signal predicts social conformity. *Neuron* **61**, 140–151 (2009).
3. D. Campbell-Meiklejohn, A. Simonsen, C. D. Frith, N. D. Daw, Independent neural computation of value from other people's confidence. *J. Neurosci.* **37**, 673–684 (2017).
4. C. Heyes, What's social about social learning? *J. Comp. Psychol.* **126**, 193–202 (2012).
5. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press Cambridge, 2018).
6. J. P. O'Doherty, P. Dayan, J. Schultz, R. Deichmann, K. Friston, R. J. Dolan, Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
7. S. Suzuki, N. Harasawa, K. Ueno, J. L. Gardner, N. Ichinohe, M. Haruno, K. Cheng, H. Nakahara, Learning to simulate others' decisions. *Neuron* **74**, 1125–1137 (2012).
8. A. Olsson, E. Knapska, B. Lindström, The neural and computational systems of social learning. *Nat. Rev. Neurosci.* **21**, 197–212 (2020).
9. O. Bartra, J. T. McGuire, J. W. Kable, The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013).
10. L. Tremblay, W. Schultz, Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
11. C. J. Burke, P. N. Tobler, M. Baddeley, W. Schultz, Neural mechanisms of observational learning. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14431–14436 (2010).
12. T. E. J. Behrens, L. T. Hunt, M. W. Woolrich, M. F. S. Rushworth, Associative learning of social value. *Nature* **456**, 245–249 (2008).
13. M. R. Hill, E. D. Boorman, I. Fried, Observational learning computations in neurons of the human anterior cingulate cortex. *Nat. Commun.* **7**, 12722 (2016).
14. S. W. C. Chang, J.-F. Gariépy, M. L. Platt, Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* **16**, 243–250 (2013).
15. F. Grabenhorst, R. Báez-Mendoza, W. Genest, G. Deco, W. Schultz, Primate amygdala neurons simulate decision processes of social partners. *Cell* **177**, 986–998.e15 (2019).
16. E. D. Boorman, J. P. O'Doherty, R. Adolphs, A. Rangel, The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron* **80**, 1558–1571 (2013).
17. C. J. Charpentier, K. Iigaya, J. P. O'Doherty, A neuro-computational account of arbitration between choice imitation and goal emulation during human observational learning. *Neuron* **106**, 687–699.e7 (2020).
18. B. De Martino, S. M. Fleming, N. Garrett, R. J. Dolan, Confidence in value-based choice. *Nat. Neurosci.* **16**, 105–110 (2012).
19. N. Persaud, P. McLeod, A. Cowey, Post-decision wagering objectively measures awareness. *Nat. Neurosci.* **10**, 257–261 (2007).
20. J. Gläscher, A. N. Hampton, J. P. O'Doherty, Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb. Cortex* **19**, 483–495 (2009).
21. R. A. Rescorla, A. R. Wagner, A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Class. Cond. II Curr. Res. Theory* **2**, 64–99 (1972).
22. B. Lau, P. W. Glimcher, Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* **84**, 555–579 (2005).
23. C. C. Ruff, E. Fehr, The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* **15**, 549–562 (2014).
24. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, 2013).
25. J. Gläscher, J. P. O'Doherty, Model-based approaches to neuroimaging: Combining reinforcement learning theory with fMRI data. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**, 501–510 (2010).
26. G. Jocham, P. M. Furlong, I. L. Kröger, M. C. Kahn, L. T. Hunt, T. E. J. Behrens, Dissociable contributions of ventromedial prefrontal and posterior parietal cortex to value-guided choice. *Neuroimage* **100**, 498–506 (2014).
27. S. Farashahi, C. H. Donahue, B. Y. Hayden, D. Lee, A. Soltani, Flexible combination of reward information across primates. *Nat. Hum. Behav.* **3**, 1215–1224 (2019).
28. K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, R. J. Dolan, Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6**, 218–229 (1997).
29. H. Barlow, The mechanical mind. *Annu. Rev. Neurosci.* **13**, 15–24 (1990).
30. R. Saxe, N. Kanwisher, People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835–1842 (2003).
31. M. Tsakiris, L. Carpenter, D. James, A. Fotopoulou, Hands only illusion: Multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects. *Exp. Brain Res.* **204**, 343–352 (2010).
32. M. A. J. Apps, M. F. S. Rushworth, S. W. C. Chang, The anterior cingulate gyrus and social cognition: Tracking the motivation of others. *Neuron* **90**, 692–707 (2016).

33. M. A. J. Apps, N. Ramnani, Contributions of the medial prefrontal cortex to social influence in economic decision-making. *Cereb. Cortex* **27**, 4635–4648 (2017).
34. A. Rangel, T. Hare, Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* **20**, 262–270 (2010).
35. R. Polanía, M. A. Nitsche, C. C. Ruff, Studying and modifying brain function with non-invasive brain stimulation. *Nat. Neurosci.* **21**, 174–187 (2018).
36. M. J. Crockett, E. Fehr, Social brains on drugs: Tools for neuromodulation in social neuroscience. *Soc. Cogn. Affect. Neurosci.* **9**, 250–254 (2014).
37. T. A. Hare, C. F. Camerer, D. T. Knoepfle, J. P. O'Doherty, A. Rangel, Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* **30**, 583–590 (2010).
38. T. E. J. Behrens, M. W. Woolrich, M. E. Walton, M. F. S. Rushworth, Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
39. C. Mathys, J. Daunizeau, K. J. Friston, K. E. Stephan, A Bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5**, 39 (2011).
40. Y. Niv, R. Daniel, A. Geana, S. J. Gershman, Y. C. Leong, A. Radulescu, R. C. Wilson, Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* **35**, 8145–8157 (2015).
41. A. Soltani, A. Izquierdo, Adaptive learning under expected and unexpected uncertainty. *Nat. Rev. Neurosci.* **20**, 635–644 (2019).
42. R. Deichmann, J. A. Gottfried, C. Hutton, R. Turner, Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* **19**, 430–441 (2003).
43. C. Hutton, A. Bork, O. Josephs, R. Deichmann, J. Ashburner, R. Turner, Image distortion correction in fMRI: A quantitative evaluation. *Neuroimage* **16**, 217–240 (2002).
44. J. Ashburner, A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
45. J. M. Pearce, G. Hall, A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
46. S. W. Kennerley, M. E. Walton, T. E. J. Behrens, M. J. Buckley, M. F. S. Rushworth, Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* **9**, 940–947 (2006).
47. J. A. Mumford, J.-B. Poline, R. A. Poldrack, Orthogonalization of regressors in fMRI models. *PLOS ONE* **10**, e0126255 (2015).
48. S. M. Smith, T. E. Nichols, Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
49. J. Gläscher, Visualization of group inference data in functional neuroimaging. *Neuroinformatics* **7**, 73–82 (2009).
50. B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, A. Riddell, Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
51. W.-Y. Ahn, N. Haines, L. Zhang, Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Comput. Psychiatr.* **1**, 24–57 (2017).
52. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
53. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2016).
54. Y. Yao, A. Vehtari, D. Simpson, A. Gelman, Using stacking to average Bayesian predictive distributions (with Discussion). *Bayesian Anal.* **13**, 917–1007 (2018).
55. S. Palminteri, V. Wyart, E. Koechlin, The importance of falsification in computational cognitive modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).
56. L. Zhang, L. Lengersdorff, N. Mikus, J. Gläscher, C. Lamm, Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Soc. Cogn. Affect. Neurosci.* **15**, 695–707 (2020).

Acknowledgments: We thank A. Bert, K. Weisel, J. Spilcke-Liss, J. Majewski, and all radiographers for help with data acquisition; N. Daw for help in developing the computational models; and C. Büchel for helpful feedback on earlier versions of the manuscript, as well as the anonymous peer reviewers who greatly improved the manuscript. **Funding:** L.Z. was supported by the International Research Training Groups "CINACS" (DFG GRK 1247), the Research Promotion Fund (FFM) for young scientists of the University Medical Center Hamburg-Eppendorf, the Vienna Science and Technology Fund (WWTF VRG13-007), and the National Natural Science Foundation of China (NSFC 71801110). J.G. was supported by the Bernstein Award for Computational Neuroscience (BMBF 01GQ1006), the Collaborative Research Center "Cross-modal learning" (DFG TRR 169), and the Collaborative Research in Computational Neuroscience (CRCNS) grant (BMBF 01GQ1603). **Author contributions:** J.G. conceived the initial research idea and supervised the project. L.Z. performed behavioral pilot testing and acquired data. L.Z. and J.G. designed and programmed final experiments, designed computational models, performed analyses, interpreted the results, and wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Preprocessed behavioral data, fMRI BOLD time series data, and custom code to perform analyses can be accessed at the GitHub repository: https://github.com/lei-zhang/SIT. Additional data related to this paper may be requested from the authors.

Submitted 26 February 2020
Accepted 7 July 2020
Published 19 August 2020
10.1126/sciadv.abb4159

Citation: L. Zhang, J. Gläscher, A brain network supporting social influences in human decision-making. *Sci. Adv.* **6**, eabb4159 (2020).