Title: A network supporting social influences in human decision-making

Authors: Lei Zhang[1,2,3]*, Jan P. Gläscher[1]*

[1]Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany.

[2]Neuropsychopharmacology and Biopsychology Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, 1010 Vienna, Austria

[3]Lead contact

*Correspondence: lei.zhang@uke.de (L.Z.) or glaescher@uke.de (J.G.).

1

Summary:

Social Decision-making is driven by normative influence (leading to public compliance) and informational influence (overwriting private beliefs), but how the brain encodes these modulating forces in probabilistic environments remains unanswered. Using a novel goal-directed learning paradigm in 185 participants, we observed opposite effects of group consensus on choice and confidence: people succumbed to the group when confronted with dissenting information, but increased their confidence when observing confirming information. Leveraging computational modeling and functional neuroimaging we captured the nuanced distinction between normative and informational influence, and identified their unique but interacting neural representations in the right temporoparietal junction (processing social information) and in prefrontal cortices (representing value computations), whose functional coupling instantiates a reward prediction error and a novel social prediction error that modulate behavioral adjustment. These results suggest that a closed-loop network between the brain's reward hub and social hub supports social influence in human decision-making.

Keywords:

Social influence, reinforcement learning, social learning, prediction error, decision neuroscience, model-based fMRI, computational modeling, hierarchical Bayesian analysis

INTRODUCTION

Most of our everyday decisions are made in a social context. This affects both big and small decisions alike: 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 in order to obtain some guidance for our own menu selection. Behavioral studies have examined social influence as expressed by conformity (Asch, 1956) and have classified two major sources of social influence: normative and informational influence (Cialdini and Goldstein, 2004; Toelch and Dolan, 2015; Fehr and Schurtenberger, 2018). Normative influence leads to public compliance, but individuals may maintain private beliefs, whereas informational influence hypothesizes that social information is integrated into the own valuation process. Neuroscience studies have recently attempted to assess the neurobiological underpinnings of both two types of influence (Klucharev et al., 2009; Campbell-Meiklejohn et al., 2010; Edelson et al., 2011; Zaki et al., 2011; Izuma and Adolphs, 2013; Campbell-Meiklejohn et al., 2017; De Martino et al., 2017; Park et al., 2017). However, results are controversial (Toelch and Dolan, 2015), and more importantly, none of them have addressed the neurocomputational distinction and interaction between normative and informational influence in conjunction with individuals' own valuation processes. This is largely due to the challenge that most studies (Klucharev et al., 2009; Campbell-Meiklejohn et al., 2010; Zaki et al., 2011; Izuma and Adolphs, 2013) relied on preference judgment tasks where no feedback was given, which hindered the investigation of private belief, and due to a lack of a comprehensive computational model that quantifies and isolates latent determinants relevant for behavioral change. Furthermore, confidence is also crucial alongside individuals' actions in decision-making (De Martino et al., 2012), however, only a few studies have examined both action and confidence when social influence is presented (Campbell-Meiklejohn et al., 2017; De Martino et al., 2017; Park et al., 2017).

Here we establish a comprehensive account of social influence in decision-making at the behavioral, computational, and neurobiological level identifying distinct, yet interacting brain regions instantiating social decision-making in humans. We ask whether social influence has a distinct neurocomputational representation, and how it is integrated with an individual's own value computation. To test this, we measured behavioral performance of learning, in combination with computational modeling and functional magnetic resonance imaging (fMRI).

3

Computational models, especially models rooted in reinforcement learning (Sutton and Barto, 1998), offers a generative framework for approximating the hidden decision processes underlying decision-making, and hence have brought considerable advances in studying the neurocomputational mechanisms (e.g., Daw et al., 2006; Gläscher et al., 2010; den Ouden et al., 2013). Although the specific neural circuitry that is recruited during reward learning through direct experience (Cooper et al, 2012) also contributes to decision-making in social contexts, additional brain networks dedicated to representing other people's knowledge and mental state are also required for facilitating learning in social contexts (e.g. Behrens, 2008; Hampton, 2008, Boorman, 2012). Given these findings, our computational model integrates direct learning instantiated by individuals' trial-and-error, together with observational learning instantiated by tracking the others' performance. This way, our models recapitulates crucial decision variables associated with behavioral adjustments, allowing us to directly probe the network of interacting brain regions.

We hypothesize that normative influence has its basis in mentalizing processes encoded in the right temporal-parietal junction (rTPJ) based on its functional role of representing others in relation to self (Frith and Frith, 1999; Saxe and Kanwisher, 2003; Hampton et al., 2008). Besides, we hypothesize that informational influence involves modulation of social learning signals by the anterior cingulate cortex (ACC), given its relevance to processing vicarious learning (Behrens et al., 2008; Suzuki et al., 2012). In addition, we anticipate that an individual's own valuation is computed via direct reinforcement learning (RL; Sutton and Barto, 1998) encoded in the ventromedial prefrontal cortex (vmPFC; Bartra et al., 2013). We further propose an interaction of two brain networks related to processing social information (e.g., rTPJ) and to reward information (e.g., striatum), whose coupling is modulated by behavioral adjustment (Hare et al., 2010).

We tested these hypotheses by employing a novel paradigm that allows multiple players to interact with each other in real-time while engaging in a probabilistic reversal learning task (PRL; e.g., Gläscher et al., 2009). Action as well as confidence were recorded before and after receiving social information, and both action and confidence were altered by social influence. We report evidence that direct valuation is integrated with vicarious valuation resulted from informational influence to make decisions, which is instantaneously affected by normative influence. We further identify two distinct networks that separately process reward information and social information, and their functional coupling substantiates a reward prediction error and a social prediction error.

4

RESULTS

Participants (N = 185) in groups of five performed the "social influence task", of which, 39 were scanned with the fMRI scanner. The task design utilized a multi-phase paradigm, enabling us to tease apart every crucial behavior under social influence (Figure 1A). Participants began each trial with their initial choice between two abstract fractals with complementary reward probabilities, followed by their first post-decision wager (an incentivized confidence rating, referred to as "bet"; De Martino et al., 2012; Persaud et al., 2007; Dotan et al., 2018; also see Star Methods). After sequentially uncovering their peers' first decisions in order of their subjective preference, participants had the opportunity to adjust both their choice and bet. The final choice and bet were then multiplied to determine the outcome on that trial. It is worth noting that participants' actual choices were communicated to every other participant via real-time connection, thus maintaining a high ecological validity of the task. The core of this task is a probabilistic reversal learning paradigm (Gläscher et al., 2009; Figure 1B). This implementation requires participants to learn and continuously update action-outcome associations, thus creating enough uncertainty such that group decisions are likely to influence the choice and bet in the 2nd decision (i.e., inferring normative influence), and examine whether the others' learning behavior at the end of the trial was integrated into their own learning (i.e., implying informational influence; see Star Methods). These dynamically evolving group decisions also allowed us to parametrically test the effect of group consensus (Figure 1C), although participants were aware that outcomes were only dependent on their own choice and not that of the group, which prevented cooperative and competitive motives.
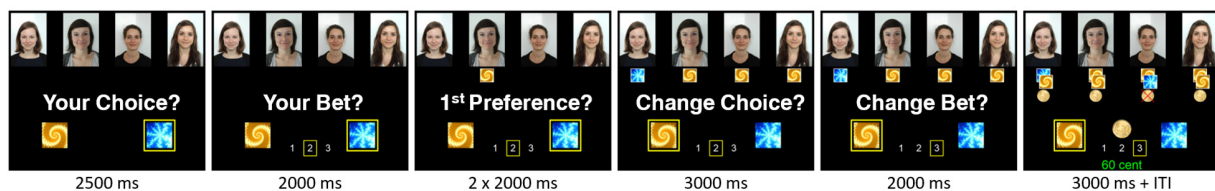
Social Influence Alters Both Action and Confidence in Goal-directed Learning

Model-free analyses showed that 185 healthy participants indeed altered both their first choice and bet after observing the group decision, but in the opposite direction. Both second choices and bets were modulated by a significant interaction between the relative direction of the group (with vs. against the participant's 1st choice) and the group consensus (2:2, 3:1, 4:0, view of each participant, Figure 1C). Participants showed an increasing trend to switch their choice toward the group decision when faced with more dissenting social information, whereas, they were more likely to persist when observing agreement with the group (direction x consensus interaction, $F_{1,574} = 55.82$, $P < 0.001$) (Figure 1D and Table S1). Conversely, participants tended to increase their bets as a
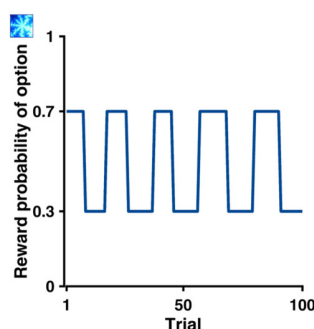
5

121 function of group consensus when observing confirming opinions, but sustained their bets when

122 being contradicted by the group ($F_{1,734} = 4.67$, $P < 0.05$) (Figure 1E and Table S1).

123   We further verified the benefit of behavior adjustment: social information facilitated

124 learning. Participants' choice accuracy of the second decision was indeed significantly higher than

125 the first one ($F_{1,2392} = 4.45$, $p < 0.05$; see Figure S1A and Table S2). Similarly, participants' second

126 bet was significantly higher than their first one ($F_{1,184} = 7.10$, $p < 0.01$; Figure S1B and Table S2).

127 Together, we identified the effect of social influence in behavioral adjustments, and demonstrated

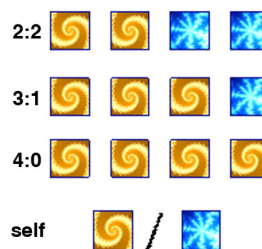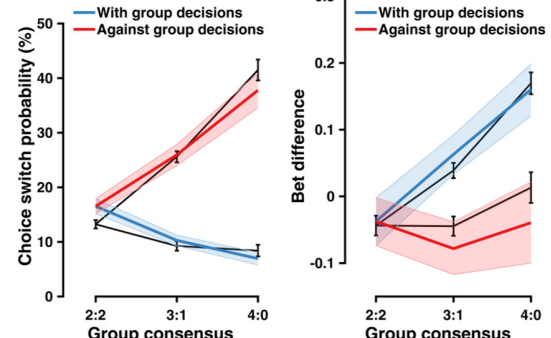128 that the adjustment is not a result of perceptual salience.

130 **Figure 1. Experimental task and behavioral results. (A)** Experimental procedure (see Star Methods).
131 **(B)** Example reward structure. Reward contingency reverses after every 8-12 trials. **(C)** Group
132 consensus (view from a participant). **(D)** Switch probability at 2nd choice and bet difference (2nd bet –
133 1st bet) as a function of the majority of the group's 1st decision (with vs. against) and the group consensus.
134 All black lines indicate actual data (mean ± within-subject SEM). Shaded error bars represent the 95%
135 highest density interval (HDI) of the mean effect computed from the winning model's posterior density
136 (posterior predictive check).

Computational mechanisms of social influence in goal-directed learning

139   Using computational modeling, we aimed to formally quantify the latent mechanisms by

140 dissociating the two types of social influence at the computational level, and particularly, by

6

unraveling how informational influence was incorporated into one's own learning process. In addition to existing RL models on social influence (Biele et al., 2011; Diaconescu et al., 2014), our model accommodates multiple players, and is able to simultaneously estimate participants' two choices and two bets under the hierarchical Bayesian analysis workflow (Gelman et al., 2013; Carpenter et al., 2017). Our efforts to construct these models were guided by two design principles: (1) separating of the individual's own value (V_{self}) and the vicarious value of others (V_{other}) during learning, which were then combined into a choice value for the 1st choice (V_{combined}) using linear weighting,

$$V_t = \beta_{\text{self}} V_{\text{self},t} + \beta_{\text{other}} V_{\text{other},t},$$

and (2) separating instantaneous normative social influence on the second choice and social learning from observing the performance of other players (i.e., informational influence). Crucially, we modeled the second choice as a function of two counteracting influences: (1) the group dissension (N_{against}) representing the instantaneous normative influence and (2) the difference between the participants' action values in the 1st choice ($V^{\text{chosen}} - V^{\text{unchosen}}$) representing the distinctiveness of the current value estimates,

$$V_t(\text{switch}) = \beta_{\text{bias}} + \beta_{\text{vdiff}}(V^{\text{chn}}_t - V^{\text{unchn}}_t) + \beta_{\text{against}} \text{w.} N_{\text{against}} + \beta_{\text{bet1}} \text{bet1}.$$

Following this construction, for instance, when the value difference on their first choice is large, participants are less likely to succumb to social influence from dissenting information on their second choice, and *vice versa*. Lastly, when all outcomes were delivered at the end of the trial, both own and vicarious value were updated on a trial-by-trial basis: V_{self} was updated with a reward prediction error (RPE; Schultz et al., 1997),

$$V_t = V_{t-1} + \eta(R_{t-1} - V_{t-1});$$

meanwhile, V_{other} was updated through tracking a preference-weighted discounted reward history (i.e., their performance in the recent past) of all four other co-players (Figure 2A; see also Star Methods),

$$V_{\text{other},t}(A) = \sum_{s=1}^{4} \sum_{i=T-3}^{T-1} \text{w}_s R_{s,i} \gamma^{T-i}, \text{ if } C2_{s,t-1} = A.$$
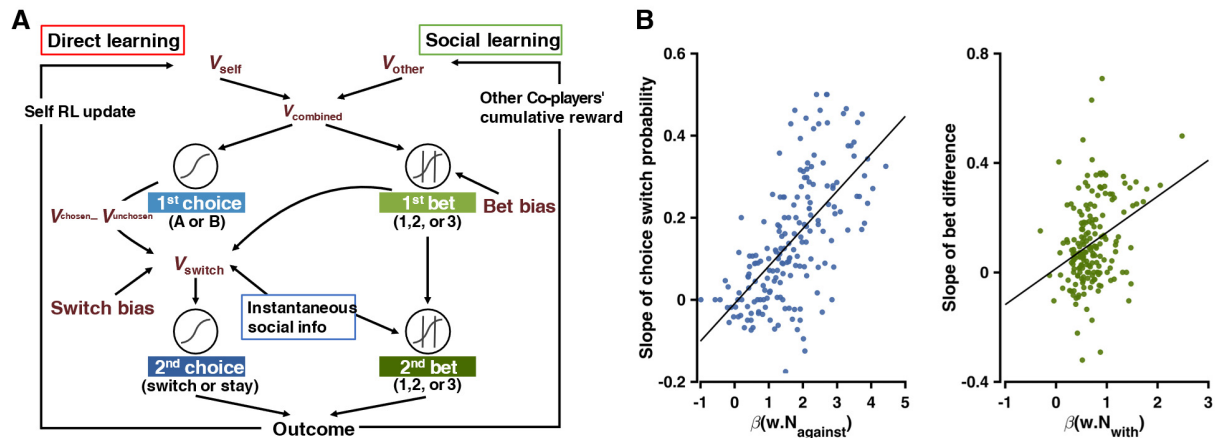
7

Figure 2. Computational model and its relation to behavior. (A) Schematic of the computational model (see main text and Star Methods). **(B)** Left: Relationship between contradicting social information (preference-weighted $N_{against}$) and the susceptibility to social influence (slope of switch probability). Right: Relationship between confirming social information (preference-weighted N_{with}) and the bet difference.

Arguably, instead of tracking each co-player's performance, individuals may simulate an RL-like algorithm to update this vicarious value through observational learning from the co-players – effectively, learning "for" the others. However, using four independent RL algorithms to update learning signals for the other four co-players is cognitively demanding – participants had to track and update each other's individual learning processes together with their own valuation to make further decisions. Given that an RL update requires both action and reward, a simpler vicarious learning mechanism may rely on either of them. In other words, participants may utilize either others' choice preference history or their performance history to approximate the value update. We tested all these hypotheses by constructing learning models with the corresponding value update rule. Model comparison first verified the necessity of the social learning component, further ruled out these alternative learning process, and therefore, confirmed that vicarious values were updated by maintaining others' discounted reward history (Table 1; see also Star Methods). Additionally, Bayesian model averaging using Bayesian bootstrap (Yao et al., 2018) indicated that the probability of this winning model being the best model over the others was 99.8%, which substantiated the model comparison result.

We further verified our winning model using two rigid validation approaches. First, we carried out a parameter recovery analysis. Although the hierarchical Bayesian approach increases

the complexity of the parameter space, all parameters of our winning model could be accurately and selectively recovered (Figure S2). Second, as model comparison provided relative model performance, we noted the importance to conduct a posterior predictive check (e.g., Frank, et al., 2015). Indeed, our winning model provided the best out-of-sample predictive power, and its posterior prediction well captured behavioral findings of our model-free analyses (Figure 1D).

Next, we sought to establish the functional association between model parameters and the model-free behaviors. Parameter results (Figure S1C-F) hinted that the extent to which participants learned from themselves and from the others was on average comparable, suggesting that an integrated value computed from one's direct learning and the informational influence to guide future decisions. Furthermore, parameters related to normative influence were well-capable of predicting the individual difference of participant's behavioral adjustment. If the model-derived signal was in high accordance with the corresponding model-free feature, we ought to anticipate a strong association between them. Indeed, we observed a significant positive correlation between $\beta(w.\text{N}_{\text{against}})$ and the slopes of choice switch probability in the against condition (Pearson's $R = 0.64$, $p < 0.001$; Figure 2B). Similarly, we observed a positive correlation between $\beta(w.\text{N}_{\text{with}})$ and slope derived from the bet difference in the "with" condition (Pearson's $R = 0.33$, $p < 0.001$; Figure 2B).

Table 1. Candidate computational models and model comparison

Class	Model	Description	ΔLOOIC
Non-social models	M1a	simple RL	0
	M1b	M1a + fictitious update	−1245
Social models with instantaneous effect	M2a	M1a + instantaneous social influence	−1893
	M2b	M1b + instantaneous social influence	−2889
Social models with instantaneous effect and observational learning	M3	M2b + OL (others' RL update)	−3089
	M4	M2b + OL (others' action preference)	−2869
	M5	M2b + OL (others' current reward)	−3203
	M6a	M2b + OL (others' cumulative reward)	−3450
	M6b	M2b + OL (others' cumulative reward) + bet1	−3507

Note: RL = reinforcement learning, OL = observational learning. LOOIC = leave-one-out information criterion; lower LOOIC value indicates better out-of-sample predictive accuracy. M6b (in bold) is the winning model.

9

Together, our computational modeling analyses suggested that (1) participants indeed learned both from their own valuation using an RPE to update their own values and from others by maintaining the others' reward history that was subsequently integrated it into their own decision process; (2) participants' behavioral adjustment was instantaneously affected by the group consensus: the number of co-players who made the opposite choice prompted participants to switch their choice towards the direction of the group, where the number of co-players who decided on the same option drove participants to increase their bet. Once we had uncovered those latent elements of the decision processes under social influence, we were then able to test how they were computed and implemented at the neural level using model-based fMRI (Gläscher and O'Doherty, 2010).

Neural substrates of dissociable self- and other value

The first part of our imaging analyses focused on how distinctive decision variables (Figure 2A) were represented in the brain (GLM1; see Star Methods). Second-level results were obtained using non-parametric methods with threshold-free cluster enhancement (TFCE; Smith and Nichols, 2009). Our model distinguished between two value signals and suggested that an integrated value signal was associated with participants' initial action and bet. Consequently, we now aimed to test the hypothesis that distinct and dissociable brain regions were recruited to implement these computational signals. Indeed, we observed that the vmPFC (peak: x = 4, y = 46, z = −14; all coordinates reported in the MNI space) activity was positively scaled with V_{self}, and the ACC (peak: x = 2, y = 10, z = 36) activity was positively scaled with V_{other} (Figure 3A; Table S3). To test whether the two value signals (i.e., V_{self}, V_{other}) are distinctively and respectively associated with vmPFC and ACC, we engaged in a double-dissociation approach (e.g., Shamay-Tsoory et al., 2009; Kennerley et al., 2011), and we found that V_{self} was exclusively encoded in the vmPFC but not in the ACC, whereas V_{other} was exclusively represented in the ACC but not in the vmPFC (Figure 3B). In addition, the medial prefrontal cortex (mPFC; peak: x = 10, y = 40, z = 10) was functionally coupled with both vmPFC and ACC (Figure S5 and Table S5), suggesting a neural encoding for the integrated value signal (e.g., Rouault et al., 2019). Besides the value signals, an RPE signal was firmly associated with activities in the nucleus accumbens (NAcc; left peak: x = −10, y = 8, z = −10; right peak: x = 12, y = 10, z = −12; Figure 3D; Table S3), a region

243   that is well-studied in the literature (e.g., Schultz et al., 1997). However, a closer look at the two

244   theoretical subcomponents of RPE is necessary to assess its neural substrates (e.g., Behrens et al.,

245   2008; Jocham et al., 2014). To qualify as a region encoding an RPE signal, activities in the NAcc

246   ought to covary positively with the actual outcome (i.e., reward) and negatively with the

247   expectation (i.e., value). Notably, this property thus provides a common framework to test the

248   neural correlates of any error-like signal (Behrens et al., 2008). Under this framework, we indeed

249   found that the activities in the NAcc showed a positive correlation with the reward outcome ($p <$

250   0.0001, permutation test; Figure 3E, green line), and a negative effect of the value signal ($p =$

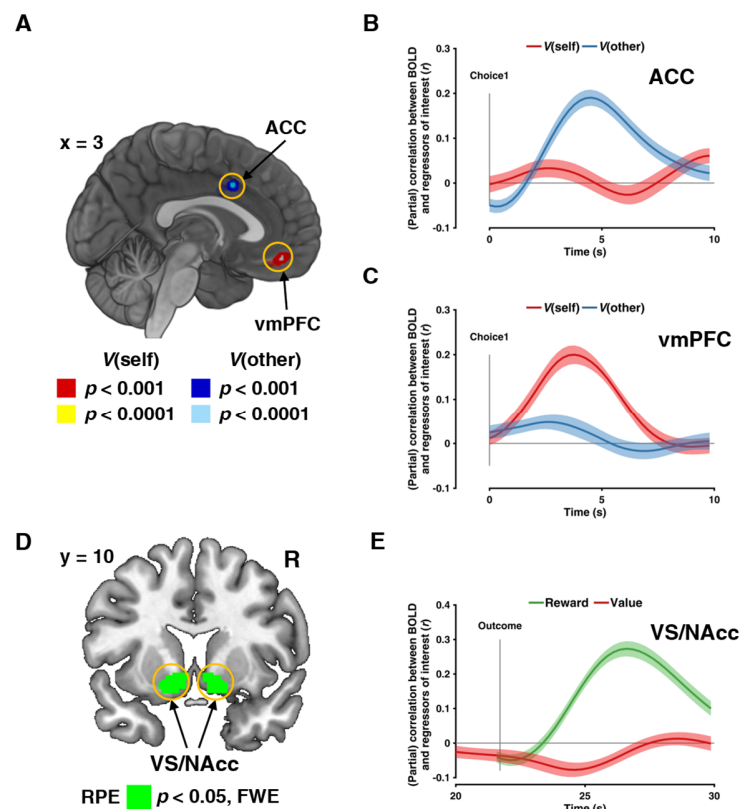251   0.021, permutation test; Figure 3E, red line).

252

253 **Figure 3. Neural substrates of dissociable value signals and reward prediction error. (A)** The neural
254   representation of V_{self} and V_{other} are encoded in the vmPFC (red/yellow) and the ACC (blue/light blue),
255   respectively. Sagittal slice at x = 3. **(B, C)** Time series estimates (Behrens et al., 2008) demonstrate a
256   double dissociation of the neural signatures of the value signals. The vmPFC is positively correlated
257   with V_{self}, but not with V_{other}, whereas the ACC is positively correlated with V_{other}, but not with V_{self}. **(D)**
258   The neural representation of reward prediction error (RPE) is encoded in the VS/NAcc. Coronal slice at
259   y = 10. **(E)** The time series in the left VS/NAcc is sensitive to both component of the RPE: positive
260   correlation with reward (green line), and negative correlation with value (red line).

261

11

Neural correlates of dissenting social information and behavioral adjustment

We next turned to disentangle the neural substrates of the instantaneous social influence (GLM1; see Star Methods) and the subsequent behavioral adjustment (GLM2; see Star Methods). As we have validated enhanced learning using such social information (Figure S1), we reasoned that participants might process other co-players' intentions relative to their own first decision to make subsequent adjustments, and this might be related to the mentalizing network. Based on this reasoning, we assessed the parametric modulation of preference-weighted normative influence ($w.N_{against}$), and indeed found that activity in TPJ (left peak: x = −48, y = −62, z = 30; right peak: x = 50, y = −60, z = 34), among other regions (e.g., ACC, anterior insula; see Table S4), was positively correlated with the dissenting social information (Figure S3; Table S4). In addition, the resulting choice adjustment (i.e., switch vs. stay) covaried with activity in bilateral dorsolateral prefrontal cortex (dlPFC; left peak: x = −32, y = 48, z = 16; right peak: x = 26, y = 42, z = 32; Figure S4; Table S4), commonly associated with executive control and behavioral flexibility (Gläscher et al., 2009; Burke et al., 2010). In contrast, the vmPFC (peak: x = 6, y = 44, z = −16) was more active during stay trials (i.e., stay > switch) trials, reminiscent of its representation of one's own valuation (Bartra et al., 2013; Gläscher et al., 2009; Figure S4; Table S4). In summary, our model-based fMRI analyses (Gläscher and O'Doherty, 2010; Cohen et al., 2017) revealed two distinct brain networks representing social information and reward and value processing.

A network between Brain's Reward Circuit and Social Circuit

So far, we have shown how key decision variables are implemented at distinct nodes at the neural level. In a next step, we sought to establish how these network nodes are functionally connected to bring about socially-induced behavioral change and to uncover additional latent computational signals that would otherwise be undetectable by conventional general linear models. We first conducted a psycho-physiological interaction (PPI, Friston et al., 1997; O'Reilly et al., 2012) to examine the context-dependent connectivity, and then we performed a physio-physiological interaction (PhiPI; Friston et al., 1997) to further interrogate the functional coupling at the physiological level.
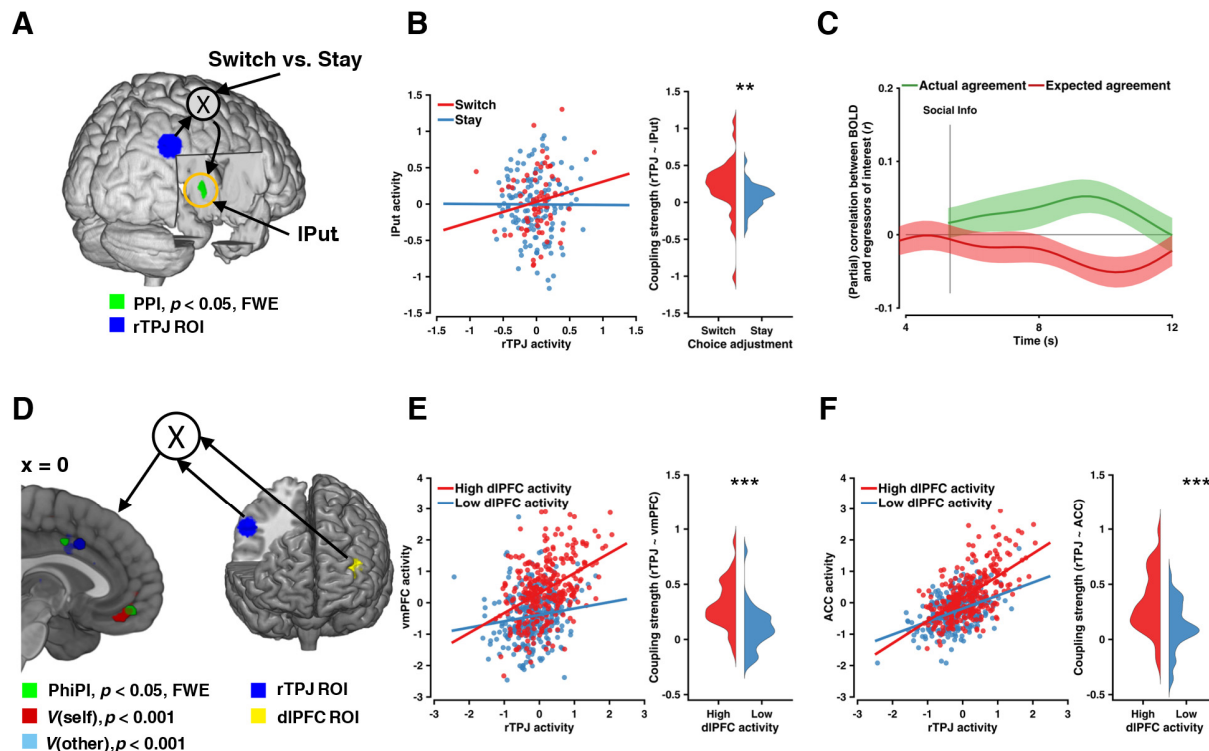
Figure 4. Functional connectivity between reward-related regions and social-related regions. (A) The functional connectivity between the left putamen (green) and the seed region rTPJ (blue) is modulated by the choice adjustment (switch vs. stay). **(B)** Correlation of activity in seed and target region for both switch and stay trials in an example subject and histogram of coupling strength across all participants for switch and stay trials. **(C)** The BOLD time series in the left putamen (PPI target) exhibits a social prediction error (positive correlation with the actual agreement, and negative correlation with the expected agreement) (mean effect across participants ± SEM). **(D)** Two seed regions, the rTPJ (blue), which responds to the social information, and the left dlPFC (yellow), which encodes the choice adjustment, elicit connectivity activations in the vmPFC and the pMFC (both in green), which partially overlap with the latent value signals (i.e., V_{self}; red; and V_{other}; blue), as in Figure 3A. Sagittal slice at x = 0. **(E, F)** Correlation plots of seed and target regions for both high and low dlPFC activity in an example subject and histograms of seed-target coupling strengths across all participants for high and low dlPFC activity.

Using a PPI seeded in rTPJ (see Star Methods), we investigated how behavioral change at the 2nd decision modulated the functional coupling between the social information represented in rTPJ and other brain regions. This analysis identified the left putamen (lPut; peak: x = −20, y = 12, z = −4; Figure 4A and 4B; Table S5). Closer investigations into the computational role of lPut revealed that it did not correlate with the two components of an RPE: activities in the lPut only positively correlated with reward ($p < 0.0001$, permutation test), but not negatively correlated with

value, (p = 0.4854, permutation test). Instead, as the choice adjustment resulted from social information, we reasoned that lPut might encode a social prediction error (SPE). Follwing this reasoning, we conducted a similar analysis as we did for the RPE, and we found that activity in the lPut was positively correlation with the actual agreement (approximated by 1-$N_{against}$%; p = 0.040, permutation test) and negatively correlated with the expected agreement (approximated by the value difference V^{chosen} - $V^{unchosen}$; as individuals who maintain a larger value difference may expect more agreement; Zhu et al., 2012; p = 0.014, permutation test) (Figure 4C). This pattern confirmed that lPut was effectively encoding a hitherto uncharacterized social prediction error. Taken together, these analyses demonstrate that functional coupling between neural representations of social information and an SPE is enhanced, when this social information is leading to a behavioral change.

In the last step, using a PhiPI we investigated how the neural representation of switching at 2nd decision in the left dlPFC modulated the functional coupling of rTPJ and other brain regions. This analysis revealed that activity in rTPJ positively modulated the coupling between vmPFC (peak: x = 0, y = 48, z = −12) and ACC (peak: x = 0, y = 0, z = 40), which strikingly overlapped with the regions that represented the two value signals (Figure 4D-F, Table S5). Therefore, it seems that the interplay of neural representations of social information and the propensity for behavioral change leads to the updating of both values signals obtained via both direct learning and observational learning.

DISCUSSION

Social influence is a powerful modulator of individual choices (Ruff and Fehr, 2014). Although accumulating studies have investigated the neural representations of social influence and attempted to identify potential mechanisms, there is little direct evidence for the dissociation between normative influence and informational influence and how the distinct computations are represented in the brain and how these brain regions interact with one another. Here, we addressed this gap with a novel social decision-making paradigm that allowed us to dissociate the two sources of social influence. In a comprehensive neurocomputational approach to social decision-making, we were not only able to identify a network of brain regions that represents and integrates social information of others, but also characterize the computational role of each node in this network in

14

detail (Figure 5), suggesting the following process model: one's own decision is guided by a combination of value signals from direct learning (V_{self}) represented in vmPFC (Figure 3A-B; Bartra et al., 2013) and from observational learning (V_{other}) represented in a section of ACC (Figure 3A-B) that is also closely related to estimates of the volatility of others' choices (Behrens et al., 2008) and to error detection and response conflict resolution (e.g., Carter, 1998). The decisions of others are encoded with respect to the own choice in rTPJ (Figure S4), an area linked, but not limited to representations of social information and agents in a variety of tasks (Saxe and Kanwisher, 2003; Hampton et al., 2008; Suzuki et al., 2015). In fact, rTPJ is also related to Theory of Mind (Frith and Frith, 1999) and other integrative computations such as multisensory integration (Tsakiris et al., 2010) and attentional processing (Corbetta and Shulman, 2002). Moreover, dissenting social information gives rise to a novel and hitherto uncharacterized social prediction error (difference between actual and expected agreement with group decision) represented in lPut (Figure 4A, 4C; Figure S5), unlike the more medial NAcc, which exhibits the neural signature of a classic RPE (Figure 3D-E; O'Doherty et al., 2003; O'Doherty et al., 2004). Notably, the interplay of lPut and rTPJ affects behavioral change toward the group decision (Figure 4A-B) in combination with its neural representation of choice switching in left dlPFC (Figure 4D-F). These functionally connected neural activities trigger the update of direct learning in vmPFC (V_{self}) and observational learning in ACC (V_{other}), thus closing the loop of decision-related computations in social contexts.
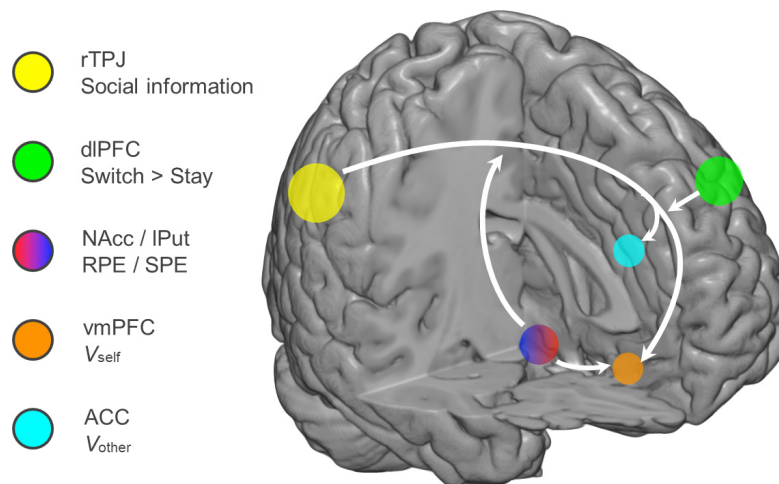

Figure 5. Schematic of the of the network supporting social influence in decision-making as uncovered in this study (for details see main text).

Our result that self-valuation is encoded in vmPFC is firmly in line with previous evidence from learning and decision-making in non-social contexts (Plassmann et al., 2007; Levy and Glimcher, 2012; Bartra et al., 2013), and extends it into a social context. On top of individuals' own value update, we further show that ACC is responding to the value signals updated from observational learning, which is aligned with previous studies that have implicated the ACC in tracking the volatility of social information (Behrens et al., 2008; Behrens et al., 2009). In particular, given that the social information in the current study is represented by the cumulative reward history of the others as inferred by our computational model, the dynamics of how well the others were performing in the recent past somewhat reflects their volatility in the same learning environment as in Behrens et al. (2008). Moreover, this distinct neural coding of direct values and vicarious values in the current study fundamentally differs from previous studies on social influence in decision-making. In a recent study, for instance, Apps and Ramnani (2017) reported that neural activities in vmPFC and ACC were respectively associated with subjective values and normative values in an intertemporal economic game. It should be noted that participants in this study were asked to separately and explicitly make intertemporal decisions either for themselves or for another group. In the current study, however, because the two value signals were modeled at the same time point, and no instruction was given to track self and other differently, we argue that the learning processes from one's own valuation and from the others' reward history were implemented in parallel; let alone our winning model has indicated the extent to which individuals were relying on their own and the others were effectively comparable (Figure S1C). Collectively, these results demonstrate concurrent yet distinct value computations in vmPFC and ACC when social information is presented during goal-directed learning.

Apart from the value dissociation, we were interested in how direct value and vicarious value were integrated to guide future decisions. As shown by our functional connectivity analyses, the mPFC covaried with activities in both the vmPFC and the ACC. According to a recent meta-analysis (Bartra et al., 2013), this region is particularly engaged during the decision stage when individuals are representing the choice options and selecting actions, especially in value-based and goal-directed decision-making (Rangel and Hare, 2010). Hence, it suggests that beyond the dissociable neural underpinnings, the directed value and vicarious value are further combined to make subsequent decisions (e.g., Rouault et al., 2019).

16

Furthermore, we replicated previous reports that identified the NAcc was associated with the RPE computation instead of mere outcome representation (Behrens et al., 2008; Jocham et al., 2014; Klein et al., 2017). That is, if a brain region encodes the RPE, its activity should be positively correlated with the actual outcome (e.g., reward), and negatively correlated with the expected outcome (e.g., value). Using this property of the RPE signal, our data identify a hitherto uncharacterized social prediction error (SPE) encoded in a section of the putamen, resulting from a psychophysiological interaction seeded at rTPJ. This suggests that the SPE signal may trigger a re-computation of expected values and give rise to the subsequent behavioral adjustment, which is partially in line previous reports showing that an SPE was signaled by increased striatal activity (Behrens et al., 2008; Meshi et al., 2012). In addition, these functional connectivity results somewhat concur with previous reports that demonstrated the rTPJ has functional links with the reward network, of which the striatal region is a central hub (Hare et al., 2010).

In addition, our results complement and extend previous neuroimaging work on social influence. Consistent with the large body of studies on social influence and conformity (Klucharev, et al., 2009; Berns et al., 2010; Tomlin et al., 2013), the ACC and the aINS are more activated when observing conflicting social information, with the ACC being relevant to general error monitoring and conflict detecting (Ridderinkhof; et al., 2004; Diedrichsen et al., 2005) and the aINS being associated with affective emotion and negative arousal (Craig, 2002, 2003). This body of evidence suggests when observing the other co-players choosing the alternative option and thus contradicting an individuals' own first choice, a conflict monitoring process may be initiated, and such conflict between individuals' prior decision and the group opinion may be accompanied by increased affective arousal, such as worry and anxiety. However, such an interpretation remains speculative as we did not collect psychophysiological measures of arousal (such as skin conductance responses). Nevertheless, it should be noted that conflict monitoring is not necessarily triggered by dissenting social information; other forms of perceptual mismatch may provoke a similar neural response in the ACC and aINS. Yet in the current study, our behavioral results have shown that switching towards the direction of the group was not due to perceptual mismatch; instead, social information was utilized to facilitate learning (Figure S1-2).

It is perhaps surprising and interesting that we did not find significant neural correlates with post-decision confidence (i.e., "bet" in the current study). This might be due to the fact that events in our current design (i.e., first choice and first bet, second choice and second bet) were not

17

constructed far apart in time, such that even carefully specified GLMs are not able to capture the variance related to the bets. More importantly, bets in the current design were closely tied to the corresponding choice valuation. In other words, when individuals were sure that one option would lead to a reward, they tend to place a high bet. In fact, this relationship was well reflected by our winning model and related model parameters (Figure S1E). That said, the bet was positively correlated with value signals, thus inevitably resulting in co-linear regressors and diminishing the statistical power. These caveats aside, our results nonetheless shed light on the change in confidence after incorporating social information in decision-making, which largely extends 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 (De Martino et al., 2017; Park et al., 2017; Campbell-Meiklejohn et al., 2017).

In summary, our results provide behavioral and computational evidence that normative social influence alters individuals' actions and confidence, and informational social influence is incorporated into their own valuation processes. Moreover, we found a network of distinct, yet interacting brain regions substantiating specific computational variables. Such a network is in a prime position to process decisions of the sorts mentioned in 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.

18

STAR METHODS:

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and MRI data should be directed to and will be fulfilled by the Lead Contact, Jan P. Gläscher (glaescher@uke.de), Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Forty-one groups of five healthy, right-handed participants were invited to participate in the study. 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. Forty-one out 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 an intranet connection, while being seated in the adjacent behavioral testing room outside the scanner. Twenty participants out of 205 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 non-responders (Tomlin et al., 2013). The final sample consisted of 185 participants (95 females; mean age: 25.56 ± 3.98 years; age range: 18-37 years), and among them, 39 participants belonged to the fMRI group (20 females; mean age: 25.59 ± 3.51 years; age range: 20-37 years). All participants 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).

METHOD DETAILS

Task

Underlying probabilistic reversal learning paradigm

The task structure of our social influence task was a probabilistic reversal learning (PRL) task. In our two-alternative forced choice PRL (Figure 1B), each choice option was associated with a

19

particular reward probability (i.e., 70% and 30%). After a variable length of trials (i.e., 8-12 trials), the reward contingencies reversed, such that individuals who were undergoing this task needed to re-adapt 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 anticorrelation structure thus existed to represent the underlying reward dynamics (Gläscher et al., 2009).

We used the PRL task rather than tasks with constant reward probability (e.g., being always 70%) because the PRL task structure requires participants continuously pay attention to the reward contingency, in order 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 assures constant learning throughout the entire experiment (Figure S1A-B). In fact, one of our early pilot studies used a fixed reward probability. In this pilot, participants quickly learned the reward contingency and neglected the social information; in this version, we thus could not tease apart the contributions between reward-based influence and socially-based influence.

Breakdown of the social influence task

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

The social influence task contained six phases. *Phase 1. Initial choice (1st choice)*. Upon the presentation of two choice options using abstract fractals, participants were asked to make their 1st choice. A yellow frame was then presented to highlight the chosen option. *Phase 2. Initial bet (1st bet)*. After making the 1st choice, participants were asked to indicate how confident they were in their choice, 1 (not confident), 2 (reasonably confident) or 3 (very confident). Notably, the confidence ratings also serve as post-decision wagering metric (an incentivized confidence rating, Persaud, et al., 2007); namely, the ratings would be multiplied by their potential outcome on each trial. For instance, if a participant won on a particular trial, the reward unit (i.e., 20 cent in the current setting) was then multiplied with the bet (e.g., 2) to obtain the final outcome (20 * 2 = 40

20

cent). Therefore, the confidence rating in the current paradigm was referred to as "bet". A yellow frame was presented to highlight the chosen bet. *Phase 3. Preference giving*. Once all participants had provided their choices and bets, the choices (but not the bets) of the other co-players 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. The remaining two choices were then displayed automatically. This manipulation is essential, because, in studies of decision-making, individuals tend to assign different credibility to their social peers based on their performance (e.g., Behrens et al., 2008; Boorman et al., 2013). In this study that there were four other co-players in the same learning environment, it is likely that they had various performance levels, and therefore shall receive difference credibility. *Phase 4. Choice adjustment (2nd choice)*. When all four other choices were presented, participants were able to adjust their choices given the social information. The yellow frame was shifted accordingly to highlight the adjusted choice. *Phase 5. Bet adjustment (2nd bet)*. After the choice adjustment, participants might adjust their bet as well. Additionally, participants also observed other co-players' second choices (on top of the first choices) once they had submitted their adjusted bets. Presenting other co-players' choices after the bet adjustment rather than the choice adjustment prevented a biased bet adjustment by the additional social information. The yellow frame was shifted accordingly to highlight the adjusted bet. *Phase 6. Outcome delivery*. Finally, the outcome was determined by the combination of participants' 2nd choice and 2nd bet. Outcomes of the other co-players were also displayed, but shown only as the single reward unit (i.e., 20 cent) without multiplying their 2nd bet. This was to provide participants with sufficient yet not overwhelming information about their peer's performance.

Procedure

To ensure a complete understanding of the task procedure, this study was composed of a two-day procedure: *pre-scanning training* (Day1), and *main experiment* (Day2).

Pre-scanning training (Day1)

21

One to two days prior to the MRI scanning (Day2), five participants came to the behavioral lab to participate in the pre-scanning 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 (1) that an intranet connection was established so that they would observe real responses from the others, (2) what probabilistic reward meant by receiving examples, (3) that there was neither cooperation nor competition in this experiment, and (4) 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. Importantly, to shift the focus of the study away from social influence and conformity, we stressed the experiment as a multi-player 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 *Reward Payment* 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 co-players without their own.

The training task contained 10 trials and differed from the main experiment in two aspects. Firstly, it used a different set of stimuli than used in the main experiment to avoid any learning effect. Secondly, participants were given a longer response window to fully understand every step of the task. Specifically, each trial began with the stimuli presentation of the two choice alternatives (4000ms), followed by the 1st bet (3000ms). After the two sequential preference ratings (3000ms each), all 1st choices from the others were displayed below their corresponding photos (3000ms). Participants were then able to adjust their choice (4000ms) and their bet (3000ms). Finally, outcomes of all participants were released (3000ms), followed by a jittered inter-trial interval (ITI, 2000 – 4000ms). To help participants familiarize themselves, we orally instructed

22

564 them what to expect and what to do on each phase for the first two to three trials. The procedure
565 during Day1 lasted about one hour.

566

567 *Main experiment* (Day2)

568 On the testing day, the five participants came to the MRI building. After a recap of all the important
569 aspects of the task instruction, the fMRI participant gave the MRI consent and entered the scanner
570 to perform the main social influence task, while the remaining 4 participants were seated in the
571 same room adjacent to the scanner to perform the same task. All computers were interconnected
572 via the intranet. They were further instructed not to make any verbal or gestural communications
573 with other participants in the experiment.

574 The main experiment contained 100 trials and used a different pair of stimuli from the
575 training task. It followed the exact description detailed above (see *Breakdown of the social*
576 *influence task*). Specifically, each trial began with the stimuli presentation of the two choice
577 alternatives (2500ms), followed by the 1st bet (2000ms). After the two sequential preference ratings
578 (2000ms each), all 1st choices from the others were displayed below their corresponding photos
579 (3000ms). Participants were then able to adjust their choice (3000ms) and their bet (2000ms).
580 Finally, outcomes of all participants were released (3000ms), followed by a jittered inter-trial
581 interval (ITI, 2000 – 4000ms). The procedure during Day2 lasted about 1.5 hours.

582

583 <u>Reward payment</u>

584 All participants were compensated with a base payment of 35 Euro plus the reward they had
585 achieved during the main experiment. In the main experiment, to prevent participants from careless
586 responses on their 1st choice, they were explicitly instructed that on each trial, either the 1st choice
587 or the 2nd choice would be used to determine the final payoff. However, this did not affect the
588 outcome delivery on the screen. Namely, although on some trials participants' 1st choice was used
589 to determine their payment, only outcomes that corresponded to the 2nd choice appeared on the
590 screen. Additionally, when their total outcome was negative, no money was deducted from their
591 final payment. Overall, participants gained 4.48 ± 4.41 Euro after completing the experiment.
592 Finally, the experiment ended with an informal debriefing session.

593

Behavioral data acquisition

Stimulus presentation, MRI pulse triggering, and response recording were accomplished with Matlab R2014b (www.mathworks.com) and Cogent2000 (www.vislab.ucl.ac.uk/cogent.php). In the behavioral group (as well as during the pre-scanning training), buttons "V" and "B" on the keyboard corresponded to the left and right choice options, respectively; and "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. To avoid motor artifacts, the position of the two choices options was counterbalanced for all the participants.

603

FMRI data acquisition and pre-processing

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 x 2 x 2 mm, with 1 mm spacing between slices) acquired using a T2*-weighted echoplanar imaging (EPI) protocol (TR, 2510ms; TE, 25ms; flip angle, 40°; FOV, 216mm) 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 (Deichmann et al., 2003). Data for each participant were collected in three runs with total volumes ranging from 1210 to 1230, and the first 3 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 = 5.00ms, TE2 = 7.46ms), and a high-resolution anatomical image (voxel size, 1 x 1 x 1 mm) was acquired after the experiment using a T1-weighted MPRAGE protocol.

fMRI data preprocessing was performed using SPM12 (Statistical Parametric Mapping; Wellcome Trust Center for Neuroimaging, University College London, London, UK). After converting raw DICOM images to NIfTI 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 (Jezzard and Balaban, 1995; Andersson et al., 2001; Hutton et al.,

24

2002). Incorporating this VDM, the EPI images were then corrected for motion and spatial distortions through realignment and unwarping (Andersson et al., 2001). 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 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 (Ashburner, 2007) 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 (FWHM) was used to smooth the EPI images.

After the preprocessing, we further detected brain volumes that (1) excessively deviated from the global mean of the BOLD signals (> 1 SD), (2) showed excessive head movement (movement parameter / TR > 0.4), or (3) 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://jan-glaescher.squarespace.com/s/spike_analyzer.m) which returned indices of those detected volumes. We then constructed them as additional participant-specific nuisance regressors of no interest across all our first-level analyses. This implementation detected $3.41 \pm 4.79\%$ of all volumes. As this procedure was done per participant, the total number of regressors for each participant may differ.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral data analysis

We tested for behavioral adjustment after observing social information in Phase 3, by assessing the choice switch probability in Phase 4 (how likely participants switched to the opposite option) and the bet difference in Phase 5 (2nd bet magnitude minus the 1st bet magnitude) as a measurement of how choice and confidence were modulated by the social information. Neither group difference (fMRI vs. behavioral) nor gender difference (male vs. female) was observed for choice switch probability (group: $F_{1,914} = 0.14$, $p = 0.71$; gender: $F_{1,914} = 0.24$, $p = 0.63$) and 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 the subsequent analysis. Additionally, trials where participants did not give valid responses on

25

652 either the 1st choice or the 1st bet in time were excluded from the sample. On average, $7.9 \pm 7.3\%$
653 of the entire trials were excluded.

654   We first tested how choice switch probability (Figure 1D, left) and bet difference (Figure
655 1D, right) varied as a function of the direction of the group (with and against, with respect to each
656 participant's 1st choice) and the consensus of the group (2:2, 3:1, 4:0, view of each participant,
657 Figure 1C). To this end, we submitted the choice switch probability and the bet difference to an
658 unbalanced 2 (direction) x 3 (consensus) repeated measures ANOVAs. The unbalance was due to
659 the fact that data in the 2:2 condition could only be used once, and we grouped it into the "against"
660 condition, resulting in three consensus levels in the "against" condition and two consensus levels
661 in the "with" condition. Grouping it into the "with" condition did not alter the results. We also
662 sought to account for the random effect in this analysis. We constructed five mixed effect models
663 (Table S1) with different random effect specifications, and selected the best one for the subsequent
664 statistical analysis.

665   We then tested if there was a linear trend within each *direction* condition as a function of the
666 *group consensus*. That is, we tested whether the choice switch probability in the "against" (or
667 "with") condition showed a significant increase (or decrease) trend as the group consensus. To this
668 aim, we first dummy coded the consensus 2:2, 3:1, 4:0 as 1, 2 and 3, then performed a simple 1st-
669 order polynomial fit using the choice switch probability as a function of the newly coded
670 consensus. We concluded the linear trend when the slope term was significant. Similarly, the linear
671 trend in the bet difference was also tested as a function of the group consensus for each direction.

672   Given that participants' interest solely lay in maximizing their personal payoffs, we then
673 tested whether it was beneficial for the participants to adjust their choice after receiving the social
674 information. If so, participants were expected to perform better (choosing the "good" option more
675 often) on their 2nd choices than on their 1st choices. To this aim, we assessed the accuracy of both
676 choices (whether selecting the more rewarding option) as well as both bets' magnitude (i.e., 1, 2
677 and 3). We selected a window of three trials to perform this analysis: three trials before the reversal
678 and three trials after the reversal, with the reversal included. We then stacked the data with respect
679 to the reversal (i.e., time-lock) and averaged them per participants. Similar to the above analysis,
680 here we submitted the data to a 2 (1st accuracy / 1st bet vs. 2nd accuracy / 2nd bet) x 7 (relative trial
681 position, $-3, -2, -1, 0, +1, +2, +3$) ANOVAs with five difference random effect specifications,

682 respectively (Table S2). If the main effect of *position* was significant, we then submitted the data

683 to a post-hoc comparison with Tukey's HSD correction.

684 All repeated measures ANOVA mixed-effect models were analyzed with the "lme4"

685 package (Bates et al., 2014) in R (v3.3.1; www.r-project.org). The 1st-order polynomial fit was

686 performed with Matlab R2014b. Results were considered statistically significant at the level $p <$

687 0.05.

688

689 <u>Computational modeling</u>

690 We developed three categories of models to uncover the latent computational mechanisms (Figure

691 2A) when participants were performing the social influence task. We based all our computational

692 models on the simple reinforcement learning model (RL, Sutton and Barto, 1998), and

693 progressively add components (Table 1).

694 First, given the structure of the PRL task, we sought to evaluate whether a fictitious update

695 RL model that incorporates the anticorrelation structure (see *Underlying probabilistic reversal*

696 *learning paradigm*) outperformed the simple RL model that only updated the value of the chosen

697 option. Thus, we constructed both the simple RL model and the fictitious update RL model, and

698 both of them did not consider social information (Category 1: M1a and M1b). On top of Category

699 1 models, we then included the instantaneous social influence (reflecting the normative influence)

700 in the non-social models to construct social models (Category 2: M2a and M2b). Finally, we

701 further considered the component of observational learning (reflecting the informational influence)

702 with competing predictions (Category 3: M3, M4, M5, M6a, M6b). In all models, we

703 simultaneously estimated participants choice and bet using hierarchical Bayesian analysis. The

704 remainder of this section explains the technical details regarding the model specification.

705

706 *Choice model specifications*

707 In all models, the 1st choice was estimated using a softmax function (Sutton and Barto, 1998)

708 that converted action values into action probabilities. On trial *t*, the action probability of choosing

709 the option A (between A and B) was defined as follows:

27

$$p_t(\text{A}) = \frac{e^{V_t(\text{A})}}{e^{V_t(\text{A})} + e^{V_t(\text{B})}} = \frac{1}{1 + e^{V_t(\text{B}) - V_t(\text{A})}}.$$
(1)

For the 2nd choice, because we coded it as a "switch" (1) or a "stay" (0), it was modeled as logistic regression with a switch value ($V(\text{switch})$). On trial t, the probability of switch given the switch value was defined as follows:

$$p_t(\text{switch}) = \Phi(V_t(\text{switch})),$$
(2)

where Φ was the inverse logistic linking function:

$$\Phi(x) = \frac{1}{1 + e^{-x}}.$$
(3)

It is worth noting that, in both action probability model specifications, we did not include a commonly-used inverse softmax temperature parameter τ. This was because we explicitly constructed both the option values in the 1st choice and the switch value in the 2nd choice in a design-matrix fashion (e.g., Eq. 5; and 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 (Gelmam et al., 2013). 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 and M1b) did not consider any social information. In the simplest model (M1a), a Rescorla-Wagner model (Rescorla and Wagner, 1972) was used to model the 1st choice, with only the chosen value being updated via the RPE (δ), and the unchosen value remaining the same as the last trial.

$$\begin{aligned}
V^{\text{chn}}_t &= V^{\text{chn}}_{t-1} + \eta \delta_{t-1} \\
\delta_{t-1} &= R_{t-1} - V^{\text{chn}}_{t-1} \\
V^{\text{unchn}}_t &= V^{\text{unchn}}_{t-1}
\end{aligned}.$$
(4)

An effect weight was then multiplied by the values before being submitted to Eq. 1, as in:

$$V_t = \beta_V [V^{\text{chn}}_t \quad V^{\text{unchn}}_t].$$
(5)

Because there was no social information in M1a, the switch value of 2nd choice was comprised merely of the value difference of the 1st choice and a switch bias:

28

$$V_t(\text{switch}) = \beta_{\text{bias}} + \beta_{\text{vdiff}}(V^{\text{chn}}{}_t - V^{\text{unchn}}{}_t) \, . \tag{6}$$

In M1b we tested whether the fictitious update could improve the model performance, as the fictitious update has been successful in PRL tasks in non-social contexts (e.g., Hampton et al., 2007; Gläscher et al., 2009). In M1b, both the chosen value and the unchosen value were updated, as in:

$$
\begin{aligned}
V^{\text{chn}}{}_t &= V^{\text{chn}}{}_{t-1} + \eta \delta^{\text{chn}}{}_{t-1} \\
\delta^{\text{chn}}{}_{t-1} &= R_{t-1} - V_{t-1}{}^{\text{chn}} \\
V^{\text{unchn}}{}_t &= V^{\text{unchn}}{}_{t-1} + \eta \delta^{\text{chn}}{}_{t-1} \\
\delta^{\text{unchn}}{}_{t-1} &= - R_{t-1} - V_{t-1}{}^{\text{unchn}} \, .
\end{aligned}
\tag{7}
$$

Our Category 2 models (M2a and M2b) tested the role of instantaneous social influence on the 2nd choice, namely, whether observing choices from the other co-players in the same learning environment contributed to the choice switching. As compared with M1 (M1a and M1b), only the switch value of the 2nd choice was modified, as follows:

$$V_t(\text{switch}) = \beta_{\text{bias}} + \beta_{\text{vdiff}}(V^{\text{chn}}{}_t - V^{\text{unchn}}{}_t) + \beta_{\text{against}} \text{w.N}_{\text{against}} \, , \tag{8}$$

where w.N$_{\text{against}}$ denoted the preference-weighted number of against relative to participants' 1st choice. This is to reflect the ordering effect based on participants' preference. Note that the preference weight were fixed parameters based on each participant's preference towards the others when uncovering their choices (see Experimental design): the 1st favored co-player received a weight of 0.75, the 2nd favored co-player received a weight of 0.5, and the rest two co-players received a weight of 0.25, respectively. Of note, estimating these preference weights as free parameters would cause unidentifiable model estimate behavior, thus this is beyond the scope of this paper. Moreover, this term (w.N$_{\text{against}}$) was normalized to lie between 0 and 1 before entering Eq. 8. All other specifications of M2a and M2b were identical to M1a and M1b, respectively.

Next, we assessed whether participants learned from their social peers and whether they updated vicarious action values through observational learning using Category 3 models (M3, M4, M5, M6a, M6b). It is worth noting that, models belonging to Category 2 solely considered the instantaneous social effect on the 2nd choice, whereas models in Category 3 tested several competing hypotheses of the observational learning effect that may contribute to the 1st choice on the following trial, in combination with individuals' own valuation processes. In all models within

760   this category, the choice value of the 1st choice was specified by a weighted sum between V_{self}

761   updated via direct learning and V_{other} updated via observational learning:

762    $$V_t = \beta_{\text{self}} V_{\text{self},t} + \beta_{\text{other}} V_{\text{other},t} . \tag{9}$$

763   M3 tested whether individuals recruited a similar RL algorithm to their own, and therefore

764   constructed the other co-players as independent RL agents to update their action values

765   respectively. To be specific, in the principle of M3, participants were assumed to update values

766   "for" the others using the fictitious update as described above (Eq. 7), and others' action values on

767   each C1 were determined by a preference-weighted sum between one's own value updated via

768   direct learning (V_{self}) and the vicarious value updated through the observational learning (V_{other}).

769   That is, observing the performance of the other group members was also influencing the learning

770   (i.e., updating) of expected values from trial-to-trial. The values of each choice option from each

771   co-player were weighted (by the preference weight w) and then summed to formulate V_{other}, as

772   follows:

773    $$V_{\text{other},t}(A) = \sum_{s=1}^{4} w_s V_{s,t-1}(A) , \tag{10}$$

774   where s denoted the index of the four other co-players. V_{other} was afterward normalized to lie

775   between -1 and 1, using Eq. 3:

776    $$V_{\text{other},t}(A) = 2\Phi(V_{\text{other},t}(A)) - 1 . \tag{11}$$

777   This normalization was to ensure that the numerical magnitude of V_{other} was comparable to

778   V_{self}, and it, therefore, made better sense to compare the size of the corresponding value-related

779   parameters (β_{self} and β_{other} in Eq. 3).

780   One may argue that having four independent RL agents as in M3 was cognitively demanding:

781   in order to accomplish so, participants had to track and update each other's individual learning

782   processes together with their own valuation. We, therefore, constructed three additional models

783   that employed simpler but distinct valuation pathways to update values through observational

784   learning. In essence, M3 considered both choices and outcomes to determine the action value. We

785   then asked if using either choices or outcomes alone may perform as well as, or even better than,

786   M3. Following this assumption, M4 updated V_{other} using only the others' action preference,

whereas M5 considered the others' current outcome to resemble the value update via observational learning.

In M4, other players' action preference is derived from the choice sequence over the last three trials using the cumulative distribution function of the beta distribution at the value of 0.5. For instance, if one co-player chose option A twice and option B once in the last three trials, then the action preference of choosing A for him/her was: betacdf (0.5, frequency of B + 1, frequency of A + 1) = betacdf (0.5, 1 + 1, 2 + 1) = 0.6875. Those action preferences (ρ) were then used to update V_{other}:

$$V_{\text{other},t}(A) = \sum_{s=1}^{4} w_s \rho_{s,t-1}, \text{ if } C2_{s,t-1} = A \quad , \tag{12}$$

where C2 denoted the 2nd choice. Note that, in this specification, only when $C2_{s,t-1}=A$, the action preference $\rho_{s,t-1}$ was used to update $V_{\text{other}}(A)$. $V_{\text{other}}(B)$ was updated in the same fashion. The values were then normalized using Eq. 11.

Likewise, M5 tested whether participants updated V_{other} using only each other's reward (R):

$$V_{\text{other},t}(A) = \sum_{s=1}^{4} w_s R_{s,t-1}, \text{ if } C2_{s,t-1} = A \quad . \tag{13}$$

These values were then normalized using Eq. 11.

Moreover, we did not rule out the possibility that participants maintained a cumulated reward history over the last a few trials instead of monitoring only the most recent outcome of the others. In fact, 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 non-social contexts (e.g., Kennerley et al., 2006; Scholl et al., 2017). By testing four lengths of trial windows (3, 4, or 5) and using a nested model comparison, we decided on a window of three past trials to accumulate other co-players' performance, and constructed such a model as M6a:

$$V_{\text{other},t}(A) = \sum_{s=1}^{4} \sum_{i=T-3}^{T-1} w_s R_{s,i} \gamma^{T-i}, \text{ if } C2_{s,t-1} = A \quad , \tag{14}$$

where i denoted the trial index from T−3 to T−1, and γ denoted the decay factor. The values were then normalized using Eq. 11.

31

812   Lastly, given that M6a was the winning model among all the models above (M1 – M6a)

813   indicated by model comparison (see below Model selection and posterior predictive check), we

814   assessed in M6b whether the 1st bet contributed to the choice switching on the 2nd choice as well,

815   as follows:

816
$$V_t(\text{switch}) = \beta_{\text{bias}} + \beta_{\text{vdiff}}(V^{\text{chn}}_t - V^{\text{unchn}}_t) + \beta_{\text{against}}\text{w.N}_{\text{against}} + \beta_{\text{bet1}}\text{bet1} \quad (15)$$

817

818   *Bet model specifications*

819   In all models, both the 1st bet and 2nd bet were modeled as an ordered-logistic regression that is

820   often used for quantifying discrete variables, like Likert-scale questionnaire data (Greene, 2003;

821   Greene and Hensher, 2010). We applied the ordered-logistic model because the bets in our study

822   indeed inferred an ordering effect. Namely, betting on 3 was higher than betting on 2, and betting

823   on 2 was higher than betting on 1. However, the difference between the bets 3 and 1 (i.e., a

824   difference of 2) was not necessarily twice as the difference between the bets 3 and 2 (i.e., a

825   difference of 1). Hence, we needed to model the distance (decision boundary) between them.

826   Moreover, despite the fact that the bets in our study could only be 1, 2, or 3, we hypothesized a

827   continuous mental process when individuals were placing bets, which satisfied the general

828   assumption of the ordered-logistic regression model (Greene, 2003).

829   There were two key components in the ordered-logistic model, the continuous utility U, and

830   the set of thresholds θ. As discussed above, we hypothesized a continuous strength of bet utility,

831   U_{bet}, which varied between the thresholds to predict the bets. In addition, a set of $K-1$ thresholds

832     $(\theta_{1, 2, ..., K-1})$ was introduced to quantify the decision boundaries, where K was the level of the

833   discrete categories. As there were three bet levels ($K = 3$), we introduced two decision thresholds,

834   θ_1 and θ_2, $(\theta_2 > \theta_1)$. As such, the predicted bets (bêt) on trial t were then represented as follows:

835
$$\hat{\text{bet}}_{i,t} = \begin{cases} 1, \text{if} -\infty < U_{\text{bet}_{i,t}} < \theta_1 \\ 2, \text{if}\ \theta_1 < U_{\text{bet}_{i,t}} < \theta_2 \quad , i = 1,2 \\ 3, \text{if}\ \theta_2 < U_{\text{bet}_{i,t}} < +\infty \end{cases} \quad (16)$$

836   where i indicated either the 1st bêt or the 2nd bêt. Because there were only two levels of threshold,

837   for simplicity, we set $\theta_1 = 0$, and $\theta_2 = \theta$, $(\theta > 0)$. To model the actual bets, a logistic function (Eq.

838   3) was used to obtain the action probability of each bet, as follows:

$$
\begin{cases}
p\left(\text{bet}_{i,t} = 1\right) = \Phi\left(-U_{\text{bet}_{i,t}}\right) \\
p\left(\text{bet}_{i,t} = 2\right) = \Phi\left(\theta - U_{\text{bet}_{i,t}}\right) - \Phi\left(-U_{\text{bet}_{i,t}}\right), \; i = 1, 2 \\
p\left(\text{bet}_{i,t} = 3\right) = 1 - \Phi\left(\theta - U_{\text{bet}_{i,t}}\right)
\end{cases}
. \tag{17}
$$

In our model specification of the 1st bet, the utility U_{bet1} was comprised of a bet bias and the value difference between the chosen option and the unchosen option. The rationale was that, the larger the value difference, the more confident individuals were expected to be, hence placing a higher bet. This utility U_{bet1} was kept identical across all models (M1a – M6b), as follows:

$$
U_{\text{bet1}_t} = \beta_{\text{bias}} + \beta_{\text{vdiff}}\left(V^{\text{chn}}_t - V^{\text{unchn}}_t\right) . \tag{18}
$$

Note that although the formula was the same as Eq. 6, the βs were independent of each other. To model the 2nd bet, we were interested in the bet change relative to the 1st bet. Therefore, the utility U_{bet2} was constructed on top of U_{bet1}. In all non-social models (M1a, M1b), the change term was represented by an intercept parameter, as follows:

$$
U_{\text{bet2}_t} = U_{\text{bet1}_t} + \beta_{\text{change}} . \tag{19}
$$

Moreover, in all social models (M2a – M6b), regardless of the observational learning effect, the change term was specified by the instantaneous social information, as follows:

$$
U_{\text{bet2}_t} = U_{\text{bet1}_t} + \beta_{\text{with}} \, \text{w.N}_{\text{with}} + \beta_{\text{against}} \, \text{w.N}_{\text{against}} . \tag{20}
$$

It should be noted that, however, despite the anticorrelation between w.N$_{\text{with}}$ and w.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 (Pearson's $R = 0.04$, $p > 0.05$). In fact, as shown in Figure S1F, w.N$_{\text{with}}$ predicted bet increase, whereas w.N$_{\text{against}}$ predicted bet decrease, suggesting their independent contributions to the bet change during the adjustment. Additionally, we constructed two other models using either w.N$_{\text{with}}$ or w.N$_{\text{against}}$, but the model performance was dramatically worse than including both of them (ΔLOOIC > 1000).

Model estimation with hierarchical Bayesian analysis

In all models, we simultaneously estimated both choices (C1, C2) and bets (B1, B2). Model estimations of all aforementioned candidate models were performed with hierarchical Bayesian analysis (HBA) (Gelman et al., 2013) using a newly developed statistical computing language Stan

(Carpenter et al., 2017) in R. Stan utilizes a Markov Chain Monte Carlo (MCMC) sampling scheme to perform full Bayesian inference and obtain the actual posterior distribution. We performed HBA rather than maximum likelihood estimation (MLE) because HBA provides much more stable and accurate estimates than MLE (Ahn et al., 2011). Following the approach in the "hBayesDM" package (Ahn et al., 2017), we assumed, for instance, that a generic individual-level parameter φ was drawn from a group-level normal distribution, namely, $\varphi \sim$ Normal (μ_φ, σ_φ), with μ_φ and σ_φ being the group-level mean and standard deviation, respectively. Both these group-level parameters were specified with weakly-informative priors (Gelman et al., 2013): $\mu_\varphi \sim$ Normal (0, 1) and $\sigma_\varphi \sim$ half-Cauchy (0, 5). This was to ensure that the MCMC sampler traveled over a sufficiently wide range to sample the entire parameter space. All parameters were unconstrained except for η / γ (both [0 1] constraint, with inverse probit transform) and θ (positive constraint, with exponential transform).

In HBA, all group-level parameters and individual-level parameters were simultaneously estimated through the Bayes' rule by incorporating behavioral data. We fit each candidate model with four independent MCMC chains using 1000 iterations after 1000 iterations for the initial algorithm warmup per chain, which resulted in 4000 valid posterior samples. Convergence of the MCMC chains was assessed both visually (from the trace plot) and through the Gelman-Rubin $\hat{R}$ Statistics (Gelman and Rubin, 1992). $\hat{R}$ values of all parameters were close to 1.0 (at most smaller than 1.1 in the current study), which indicated adequate convergence.

Model selection and posterior predictive check

For model comparison and model selection, we computed the Leave-One-Out information criterion (LOOIC) score per candidate model (Vehtari et al., 2016). The LOOIC score provides the point-wise estimate of out-of-sample predictive accuracy in a fully Bayesian way, which is more reliable compared to point-estimate information criterion (e.g., Akaike information criterion, AIC; deviance information criterion, DIC). By convention, lower LOOIC score indicates better out-of-sample prediction accuracy of the candidate model. Plus, a difference score of 10 on the information criterion scale is considered decisive (Burnham and Anderson, 2004). We selected the model with the lowest LOOIC as the winning model. We additionally performed Bayesian model averaging (BMA) with Bayesian bootstrap (Yao et al., 2018) to compute the probability of each

895 candidate model being the best model. Conventionally, the BMA probability of 0.9 (or higher) is
896 a decisive indication.

897 Given that model comparison provided merely relative performance among candidate
898 models (Palminteri et al., 2017), we then tested how well our winning model's posterior prediction
899 was able to replicate the key features of the observed data (a.k.a., posterior predictive check, PPC).
900 To this end, we applied a post-hoc absolute-fit approach (Steingroever et al., 2014) that factored
901 in participants' actual action and outcome sequences to generate predictions with the entire
902 posterior MCMC samples. Namely, we let the model generate choices and bets as many times as
903 the number of samples (i.e., 4000 times) per trial per participants and we asked whether the
904 generated data could reproduce the behavioral pattern in our behavioral analysis.

905 Lastly, we tested how specific model parameters linked with model-free behavior to assess
906 individual differences (Figure 2B). In the choice model, we tested the simple Pearson's correlation
907 between β(w.$N_{against}$) and the 1st-order polynomial slope derived from the choice switch probability
908 as a function of the group consensus in the "against" condition (see above Behavioral analysis).
909 Likewise, in the bet model, we tested the simple Pearson's correlation between β(w.N_{with}) and the
910 1st-order polynomial slope derived from the bet difference as a function of the group consensus in
911 the "with" condition (see above Behavioral analysis).

912

Parameter recovery

914 Considering that there were multiple free parameters in the winning model, we verified whether
915 parameters were identifiable using parameter recovery after the model fitting. In the first step, we
916 randomly drew a set of group-level parameters from the joint posterior group-distribution of M6b.
917 Next, we simulated 80 synthetic participants, whose parameters were randomly drawn from this
918 set of group-level parameters. Then, we used the model (M6b) as a generative tool to simulate
919 behavioral data for our social influence task, namely, to simulate 1st choice, 2nd choice, 1st bet, and
920 2nd bet for 100 trials per participant. Once having the behavioral data, we fit M6b to the simulated
921 data in the same way as we did for the real data. And finally, we compared whether the posterior
922 group-distribution given the simulated data recovered the actual group-level parameters that were
923 used to simulate those data (Figure S2).

35

924

925 <u>MRI data analysis</u>

926

927 *Deriving internal computational signals*

928 Based on the winning model (Table 1) and its parameter estimation (Figure S2C-F), we

929 derived the trial-by-trial computational signals for each individual MRI participant using the mean

930 of the posterior distribution of the parameters. We used the mean rather than the mode (i.e., the

931 peak) because in MCMC, especially HMC implemented in Stan, the mean is much more stable

932 than the mode to serve as the point estimate of the entire posterior distribution (Carpenter et al.,

933 2017). In fact, as we modeled all parameter as normal distributions, the posterior mean and the

934 posterior mode are highly correlated (Pearson's $R = 0.99$, $p < 0.001$).

935

936 *First-level analysis*

937 fMRI data analysis was performed using SPM12. We conducted model-based fMRI analysis

938 (Gläscher et al., 2009; Gläscher and O'Doherty, 2010) containing the computational signals

939 described above (Table S6). We set up two event-related general linear models (GLM1 and GLM2)

940 to test our hypotheses.

941 GLM1 assessed the neural representations of valuation resulted from participants' direct

942 learning and observational learning in Phase 1, as well we the instantaneous social influence in

943 Phase 3. The first-level design matrix in GLM1 consisted of constant terms, nuisance regressors

944 detected by the "Spike Analyzer", plus the following 22 regressors: 5 experimentally measured

945 onset regressors for each cue (cue of the 1st choice, cue of the 2nd choice, cue of the 1st bet, cue of

946 the 1st bet, and cue of the outcome); 6 parametric modulators (PM) of each corresponding cue

947 (V_{self}^{chosen}, V_{other}^{chosen}, belonging to the cue of the 1st choice; w. $N_{against}$ belonging to the cue of the

948 2nd choice; U_{bet1}, U_{bet2}, belonging to the cue of the 1st bet and the 2nd bet, respectively; and RPE

949 belonging to the cue of the outcome); 5 nuisance regressors accounted for all of the "no-response"

950 trials for each cue; and 6 movement parameters. Note that for the two value signals, V_{other}^{chosen} was

951 orthogonalized with respect to V_{self}^{chosen}. This allowed to obtain as much variance as possible on the

952 V_{self}^{chosen} regressor, and then any additional (explainable) variance was accounted for by the V_{other}^{chosen}

36

regressor (Mumford et al., 2015; Norbury et al., 2018). Also, we intentionally did not include the reward outcome at the outcome cue. This was because (1) the RPE and the reward outcome are known to be correlated in goal-directed learning studies using model-based fMRI (e.g., Chien et al., 2016), and (2) we sought to explicitly verify RPE signals by its hallmarks using the ROI time series extracted from each participant given the second-level RPE contrast (see below *ROI time series analysis* below).

GLM2 was set up to examine the neural correlates of choice adjustment in Phase 4. To this end, GLM2 was identical to GLM1, except that the PM regressor of w. $N_{against}$ under the cue of the 2nd choice was replaced by the PM regressor SwSt.

Second-level analysis

The resulting β images from each participant's first-level GLM were then used in a 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 the functional imaging data, we employed the threshold-free cluster enhancement (TFCE; Smith and Nichols, 2009) 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 takes the raw statistics from the second-level analysis and performs a permutation-based non-parametric test (i.e., 5000 permutations in the current study) to obtain robust results. According to previous work on the direct value signal in the vmPFC (Bartra et al., 2013) and vicarious value of the social information in the ACC (e.g., Behrens et al., 2008; Boorman et al., 2013), we performed small volume corrections (SVC) using 10-mm search volumes around the peak MNI coordinates of the vmPFC and the ACC in the corresponding studies with the TFCE correction at $p < 0.05$, FWE (family-wise error) corrected. For the otherwise whole-brain analysis, we performed whole-brain TFCE correction at $p < 0.05$, FWE corrected.

Follow-up ROI analysis

37

Depending on the hypotheses, the research question, and the corresponding PM regressors, we employed 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. For each participant, we first defined a 10-mm search volume around the peak coordinate of the second level contrast (threshold: $p < 0.001$, uncorrected); within this search volume, we then searched for each participant's individual peak and created a new 10-mm sphere around this individual peak as the ROI mask. Finally, supra-threshold voxels in the new participant-specific ROI were used for the ROI analyses.

First, 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 3 ROIs to perform the time series estimates, the vmPFC, the ACC, and the VS/NAcc.

We followed the procedure established by previous studies (Behrens et al., 2008; Jocham et al., 2014; Klein et al., 2017) 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 30s, where the cue of the 1st choice was presented at 0s, the cue of the 1st bet was presented at 2.92s, the cue of the 2nd choice was displayed at 12.82s, the cue of the 2nd bet was displayed at 16.25s, and the outcome was presented at 21.71s. All these time points corresponded to the mean onsets for each cue across trials and participants. Afterward, time series were up-sampled to a resolution of 250ms (1/10 of TR) using 2D cubic spline interpolation, resulting in a data matrix of size m x n, where m is the number of trials, and n is the number of the up-sampled time points (i.e., 30s / 250ms = 120 time points). A linear regression model containing the PMs was then estimated at each time point (across trials) for each participant. It should be noted 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 in the entire trial across all the trials. The resulting time courses of effect sizes (regression coefficients) were finally averaged across participants. Because both the time series and the PMs were normalized, these time courses of effect sizes, in fact, reflected the partial correlation between the ROI time series and PMs.

38

To test group-level significance, we employed a permutation procedure. For the time sources of effect sizes for each ROI, we defined a time window of 3-7s 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 effect sizes for 5000 repetitions to generate a null distribution, and asked 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.

Second, the Percent signal change (PSC) estimates were applied when only one PM was associated with each ROI. Particularly, we asked whether there was a linear trend of the PSC for each ROI as a function of the PM. In the current study, we defined 7 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/pMFC, the right aINS and the FPC; two ROIs were associated with the PM regressor of SwSt, being the left dlPFC and the ACC; and one ROI was associated with the inverse contrast of SwSt (i.e., StSw, stay vs. switch), being the vmPFC.

To compute the PSC, we used the "rfxplot" toolbox (Gläscher, 2009) to extract the time series from the above ROIs. The "rfxplot" toolbox further divided the corresponding PMs into different bins (e.g., 2 bins, the 1st 50% of the PM and the 2nd 50% of the PM) and computed the PSC for each bin, which resulted in a p x q PSC matrix, where p is the number of participants, and q is the number of bins. To test for significance, we performed a simple 1st-order polynomial fit using the PSC as a function of the binned PM, and asked whether the slope of this polynomial fit was significantly different from zero.

Connectivity analysis

We employed two types of connectivity analyses (Friston, et al., 1997) in the current study, the psychophysiological interaction (PPI) and the physiophysiological interaction (PhiPI) to test the functional network using fMRI (O'Reilly et al., 2012).

The psychophysiological interaction (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 seeded the entire BOLD time series from a 10-mm spherical ROI in the rTPJ, centered at the peak

1039   coordinates for w.N$_{against}$ (threshold: $p < 0.001$, uncorrected), which was detected at the onset cue

1040   of the second choice. Next, we constructed the PPI regressor by combining the rTPJ ROI signals

1041   with the SwSt variable that took place after the cue of the 2nd choice (Figure 4A-B). The first-level

1042   PPI design matrix consisted of three PPI regressors (the BOLD time series of the seed region, the

1043   modulating psychological variable, and their interaction) and all the same nuisance regressors as

1044   the above first-level GLMs. The first-level interaction regressor was then submitted to a second-

1045   level t-test to establish the group-level connectivity results, with TFCE correction $p < 0.05$, FWE

1046   corrected.

1047   The Physiophysiological interaction (PhiPI) analysis follows the same principles as the PPI

1048   analysis, except that the psychological variable in the PPI regressors is replaced by the BOLD time

1049   series from a second seed ROI. We performed two PhiPI analyses. In the first PhiPI, we used as

1050   seeds the entire BOLD time series in two 10-mm spherical ROIs in the vmPFC and the ACC, both

1051   of which were detected at the cue of the 1st choice (Figure S6). In the second PhiPI, we seeded

1052   with the entire BOLD time series from an identical 10-mm spherical ROI in the rTPJ as described

1053   in our PPI, and from a 10-mm spherical ROI in the left dlPFC, which was detected at the cue of

1054   the 2nd choice (Figure 4D-F). The setup of the first-level PhiPI design matrix and the statistical test

1055   procedure on the second-level were the same as for the PPI analysis.

1056

1057

40

DATA AND SOFTWARE AVAILABILITY

Raw behavioral data and custom code to perform analyses can be accessed on the GitHub repository: https://github.com/lei-zhang/zhang_glaescher_socialinfluence.

SUPPLEMENTARY INFORMATION:

Supplementary Information includes 6 figures and 5 tables can be found with this article at https://xxx.

ACKNOWLEDGMENTS:

AUTHOR CONTRIBUTIONS:

J.G. conceived the research idea. L.Z. and J.G. designed and programmed experiments. L.Z. acquired data. L.Z. and J.G. designed computational models. L.Z. and J.G. performed analyses and wrote the paper. J.G. supervised the project.

DECLARATION OF INTERESTS:

The authors declare no competing financial interests.

1085 **KEY RESOURCES TABLE**

1086

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw behavioral data	This paper	https://github.com/lei-zhang/zhang_glaescher_socialinfluence
fMRI data	This paper	N/A
Software and Algorithms		
Statistical Parametric Mapping 12 (SPM12)	Wellcome Trust, London	https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
TFCE Toolbox	TFCE Toolbox developers	dbm.neuro.uni-jena.de/tfce/
rfxplot	Gläscher, 2009	http://rfxplot.sourceforge.net/
MATLAB R2014b	MathWorks	https://www.mathworks.com/
Cogent 2000	Cogent 2000 developers	http://www.vislab.ucl.ac.uk/cogent_2000.php
R Version 3.3.1	R development core team	https://www.r-project.org/
LME4	Bates et al., 2014	https://cran.r-project.org/web/packages/lme4/
Stan/RStan Version 2.17.0	Carpenter et al., 2017	https://mc-stan.org/
MRIcroGL	McCausland Center for Brain Imaging	https://www.nitrc.org/projects/mricrogl
Custom code (to run analyses and produce figures)	This paper	https://github.com/lei-zhang/zhang_glaescher_socialinfluence

1087

1088

1089

REFERENCES:

Ahn, W.-Y., Haines, N., and Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. Comput. Psychiatry *1*, 24–57.

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J.R., and Brown, J.W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. J. Neurosci. Psychol. Econ. *4*, 95–110.

Andersson, J.L.R., Hutton, C., Ashburner, J., Turner, R., and Friston, K. (2001). Modeling geometric deformations in EPI time series. Neuroimage *13*, 903–919.

Apps, M.A.J., and Ramnani, N. (2017). Contributions of the Medial Prefrontal Cortex to Social Influence in Economic Decision-Making. Cereb. Cortex *27*, 4635–4648.

Asch, S.E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. Psychol. Monogr. Gen. Appl. *70*, 1–70.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. Neuroimage *38*, 95–113.

Bartra, O., McGuire, J.T., and Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage *76*, 412–427.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. Psychol. Med. *45*, 361–373.

Behrens, T.E.J., Hunt, L.T., and Rushworth, M.F.S. (2009). The Computation of Social Behavior. Science *324*, 1160–1164.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. Nature *456*, 245–249.

Berns, G.S., Capra, C.M., Moore, S., and Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. Neuroimage *49*, 2687–2696.

Boorman, E.D., O'Doherty, J.P., Adolphs, R., and Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. Neuron *80*, 1558–1571.

Burke, C.J., Tobler, P.N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. Proc. Natl. Acad. Sci. U. S. A. *107*, 14431–14436.

Burnham, K.P., and Anderson, D.R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociol. Methods Res. *33*, 261–304.

Campbell-Meiklejohn, D.K., Bach, D.R., Roepstorff, A., Dolan, R.J., and Frith, C.D. (2010). How the opinion of others affects our valuation of objects. Curr. Biol. *20*, 1165–1170.

Campbell-Meiklejohn, D., Simonsen, A., Frith, C.D., and Daw, N.D. (2017). Independent Neural Computation of Value from Other People's Confidence. J. Neurosci. *37*, 673–684.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A Probabilistic Programming Language. J. Stat. Softw. *76*.

Carter, C.S. (1998). Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance. Science *280*, 747–749.

Chien, S., Wiehler, A., Spezio, M., and Gläscher, J.P. (2016). Congruence of Inherent and Acquired Values Facilitates Reward-Based Decision-Making. J. Neurosci. *36*, 5003–5012.

Cialdini, R.B., and Goldstein, N.J. (2004). Social influence: compliance and conformity. Annu. Rev. Psychol. *55*, 591–621.

43

Cohen, J.D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K.A., Pillow, J., Ramadge, P.J., Turk-Browne, N.B., et al. (2017). Computational approaches to fMRI analysis. Nat. Neurosci. *20*, 304–313.

Cooper, J.C., Dunne, S., Furey, T., and O'Doherty, J.P. (2012). Human Dorsal Striatum Encodes Prediction Errors during Observational Learning of Instrumental Actions. J. Cogn. Neurosci. *24*, 106–118.

Corbetta, M., and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. *3*, 201–215.

Craig, A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nat. Rev. Neurosci. *3*, 655–666.

Craig, A.D. (2003). Pain mechanisms: Labeled lines versus convergence in central processing. Annu. Rev. Neurosci. *26*, 1–30.

Daw, N.D., O'Doherty, J.P., Dayan, P., Dolan, R.J., and Seymour, B. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., and Love, B.C. (2017). Social Information Is Integrated into Value and Confidence Judgments According to Its Reliability. J. Neurosci. *37*, 6066–6074.

De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2012). Confidence in value-based choice. Nat. Neurosci. *16*, 105–110.

Deichmann, R., Gottfried, J.A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage *19*, 430–441.

den Ouden, H.E.M., Daw, N.D., Fernandez, G., Elshout, J.A., Rijpkema, M., Hoogman, M., Franke, B., and Cools, R. (2013). Dissociable Effects of Dopamine and Serotonin on Reversal Learning. Neuron *80*, 1090–1100.

Diaconescu, A.O., Mathys, C., Weber, L. a. E., Daunizeau, J., Kasper, L., Lomakina, E.I., Fehr, E., and Stephan, K.E. (2014). Inferring on the Intentions of Others by Hierarchical Bayesian Learning. PLoS Comput. Biol. *10*, e1003810.

Diedrichsen, J., Hashambhoy, Y., Rane, T., and Shadmehr, R. (2005). Neural Correlates of Reach Errors. J. Neurosci. *25*, 9919–9931.

Dotan, D., Meyniel, F., and Dehaene, S. (2018). On-line confidence monitoring during decision making. Cognition *171*, 112–121.

Edelson, M., Sharot, T., Dolan, R.J., and Dudai, Y. (2011). Following the crowd: brain substrates of long-term memory conformity. Science 333, 108–111.

Fehr, E., and Schurtenberger, I. (2018). Normative foundations of human cooperation. Nat. Hum. Behav. *2*, 458–468.

Frank, M.J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J.F., and Badre, D. (2015). fMRI and EEG Predictors of Dynamic Decision Parameters during Human Reinforcement Learning. J. Neurosci. *35*, 485–494.

Frith, C.D., and Frith, U. (1999). Interacting minds--a biological basis. Science *286*, 1692–1695.

Friston, K.J., Dolan, R.J., Buechel, C., Rolls, E., Fink, G.R., and Morris, J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. Neuroimage *6*, 218–229.

Gelman, A., and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Stat. Sci. *7*, 457–472.

Gelman, A., Stern, H.S., Carlin, J.B., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). Bayesian data analysis (Chapman and Hall/CRC).

Biele, G., Rieskamp, J., Krugel, L.K., and Heekeren, H.R. (2011). The Neural basis of following advice. PLoS Biol. *9*.

Gläscher, J. (2009). Visualization of group inference data in functional neuroimaging. Neuroinformatics *7*, 73–82.

Gläscher, J.P., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron *66*, 585–595.

Gläscher, J.P., Hampton, A.N., and O'Doherty, J.P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb. Cortex *19*, 483–495.

Gläscher, J.P., and O'Doherty, J.P. (2010). Model-based approaches to neuroimaging: Combining reinforcement learning theory with fMRI data. Wiley Interdiscip. Rev. Cogn. Sci. *1*, 501–510.

Greene, W.H. (2003). Econometric analysis (Pearson Education India).

Greene, W.H., and Hensher, D.A. (2010). Modeling ordered choices: A primer (Cambridge University Press).

Hampton, A.N., Adolphs, R., Tyszka, M.J., and O'Doherty, J.P. (2007). Contributions of the Amygdala to Reward Expectancy and Choice Signals in Human Prefrontal Cortex. Neuron *55*, 545–555.

Hampton, A.N., Bossaerts, P., and O'Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc. Natl. Acad. Sci. U. S. A. *105*, 6741–6746.

Hare, T. A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., and Rangel, A. (2010). Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition. J. Neurosci. *30*, 583–590.

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., and Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. Neuroimage *16*, 217–240.

Izuma, K., and Adolphs, R. (2013). Social manipulation of preference in the human brain. Neuron *78*, 563–573.

Jezzard, P., and Balaban, R.S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. Magn. Reson. Med. *34*, 65–73.

Jocham, G., Furlong, P.M., Kröger, I.L., Kahn, M.C., Hunt, L.T., and Behrens, T.E.J. (2014). Dissociable contributions of ventromedial prefrontal and posterior parietal cortex to value-guided choice. Neuroimage *100*, 498–506.

Kennerley, S.W.S., Behrens, T.E.J., and Wallis, J.J.D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat. Neurosci. *s*, 1581–1589.

Kennerley, S.W.S., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. *9*, 940–947.

Klein, T.A., Ullsperger, M., and Jocham, G. (2017). Learning relative values in the striatum induces violations of normative decision making. Nat. Commun. *8*, 1–12.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., and Fernández, G. (2009). Reinforcement learning signal predicts social conformity. Neuron *61*, 140–151.

45

Levy, D.J., and Glimcher, P.W. (2012). The root of all value: a neural common currency for choice. Curr. Opin. Neurobiol. *22*, 1027–1038.

Meshi, D., Biele, G., Korn, C.W., and Heekeren, H.R. (2012). How Expert Advice Influences Decision Making. PLoS One *7*, 1–12.

Mumford, J.A., Poline, J.-B., and Poldrack, R.A. (2015). Orthogonalization of Regressors in fMRI Models. PLoS One *10*, e0126255.

Norbury, A., Robbins, T.W., and Seymour, B. (2018). Value generalization in human avoidance learning. eLife 7, 1–30.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. Neuron *38*, 329–337.

O'Doherty, J.P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science *304*, 452–454.

O'Reilly, J.X., Woolrich, M.W., Behrens, T.E.J., Smith, S.M., and Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. Soc. Cogn. Affect. Neurosci. *7*, 604–609.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. Trends Cogn. Sci. *21*, 425–433.

Park, S.A., Goïame, S., O'Connor, D.A., and Dreher, J.-C. (2017). Integration of individual and social information for decision-making in groups of different sizes. PLoS Biol. *15*, e2001958.

Persaud, N., McLeod, P., and Cowey, A. (2007). Post-decision wagering objectively measures awareness. Nat. Neurosci. *10*, 257–261.

Plassmann, H., O'Doherty, J., and Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J. Neurosci. *27*, 9984–9988.

Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. Curr. Opin. Neurobiol. *20*, 262–270.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Class. Cond. II Curr. Res. Theory *2*, 64–99.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443–447.

Rouault, M., Drugowitsch, J., and Koechlin, E. (2019). Prefrontal mechanisms combining rewards and beliefs in human decision-making. Nat. Commun. *10*.

Ruff, C.C., and Fehr, E. (2014). The neurobiology of rewards and values in social decision making. Nat. Rev. Neurosci. *15*, 549–562.

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." Neuroimage *19*, 1835–1842.

Shamay-Tsoory, S.G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain *132*, 617–627.

Scholl, J., Kolling, N., Nelissen, N., Stagg, C.J., Harmer, C.J., and Rushworth, M.F.S. (2017). Excitation and inhibition in anterior cingulate predict use of past experiences. eLife *6*, 1–15.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A Neural Substrate of Prediction and Reward. Science *275*, 1593–1599.

1256 Smith, S.M., and Nichols, T.E. (2009). Threshold-free cluster enhancement: Addressing problems of
1257 smoothing, threshold dependence and localisation in cluster inference. Neuroimage *44*, 83–98.

1258 Steingroever, H., Wetzels, R., and Wagenmakers, E.J. (2014). Absolute performance of reinforcement-
1259 learning models for the Iowa Gambling Task. Decision *1*, 161–183.

1260 Sutton, R.S., and Barto, A.G. (1998). Introduction to reinforcement learning (MIT press Cambridge).

1261 Suzuki, S., Adachi, R., Dunne, S., Bossaerts, P., and O'Doherty, J.P. (2015). Neural Mechanisms
1262 Underlying Human Consensus Decision-Making. Neuron *86*, 591–602.

1263 Suzuki, S., Harasawa, N., Ueno, K., Gardner, J.L., Ichinohe, N., Haruno, M., Cheng, K., and Nakahara, H.
1264 (2012). Learning to Simulate Others' Decisions. Neuron *74*, 1125–1137.

1265 Toelch, U., and Dolan, R.J. (2015). Informational and Normative Influences in Conformity from a
1266 Neurocomputational Perspective. Trends Cogn. Sci. *19*, 579–589.

1267 Tomlin, D., Nedic, A., Prentice, D. a., Holmes, P., and Cohen, J.D. (2013). The Neural Substrates of Social
1268 Influence on Decision Making. PLoS One *8*, e52630.

1269 Tsakiris, M., Carpenter, L., James, D., and Fotopoulou, A. (2010). Hands only illusion: Multisensory
1270 integration elicits sense of ownership for body parts but not for non-corporeal objects. Exp. Brain Res. *204*,
1271 343–352.

1272 Vehtari, A., Gelman, A., and Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out
1273 cross-validation and WAIC. Stat. Comput. *27*, 1–20.

1274 Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). Using Stacking to Average Bayesian Predictive
1275 Distributions (with Discussion). Bayesian Anal. *13*, 917–1007.

1276 Zaki, J., Schirmer, J., and Mitchell, J.P. (2011). Social influence modulates the neural computation of value.
1277 Psychol. Sci. *22*, 894–900.

1278 Zhu, L., Mathewson, K.E., and Hsu, M. (2012). Dissociable neural representations of reinforcement and
1279 belief prediction errors underlie strategic learning. Proc. Natl. Acad. Sci. *109*, 1419–1424.
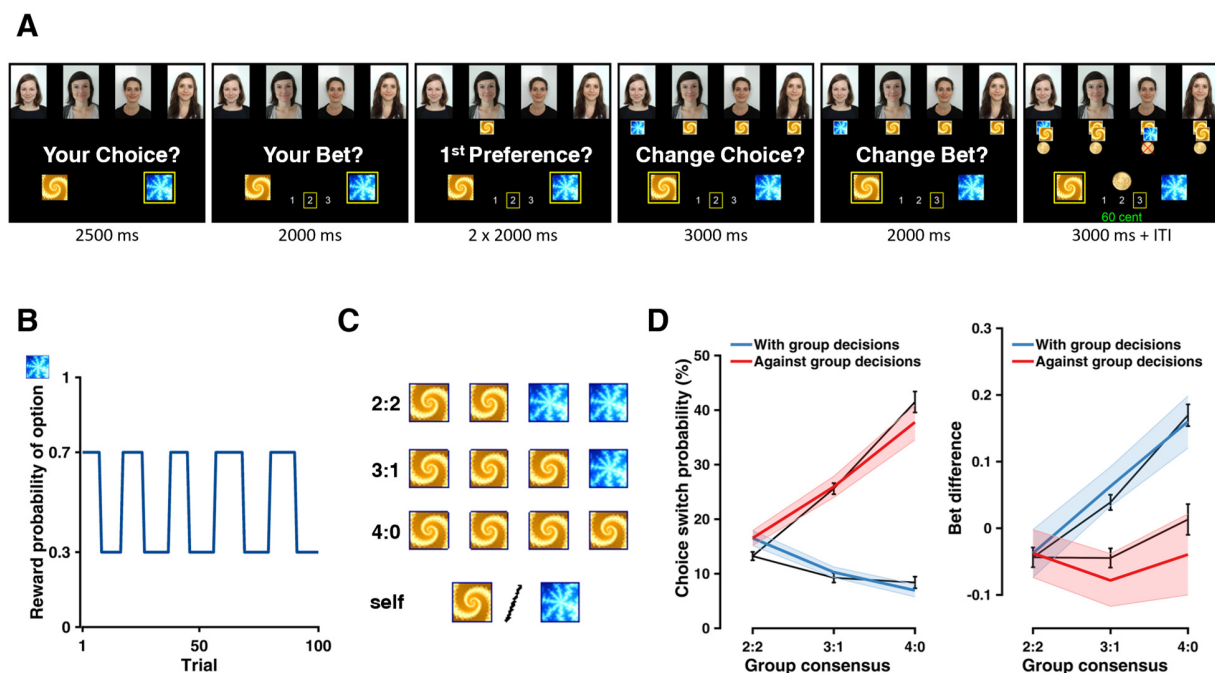1280

1281

1282

1283

Figure 1. Experimental task and behavioral results. (A) Experimental procedure (see Star Methods). **(B)** Example reward structure. Reward contingency reverses after every 8-12 trials. **(C)** Group consensus (view from a participant). **(D)** Switch probability at 2nd choice and bet difference (2nd bet – 1st bet) as a function of the majority of the group's 1st decision (with vs. against) and the group consensus. All black lines indicate actual data (mean ± within-subject SEM). Shaded error bars represent the 95% highest density interval (HDI) of the mean effect computed from the winning model's posterior density (posterior predictive check).

1291

Figure 2. Computational model and its relation to behavior. (A) Schematic of the computational model (see main text and Star Methods). **(B)** Left: Relationship between contradicting social information (preference-weighted $N_{against}$) and the susceptibility to social influence (slope of switch probability). Right: Relationship between confirming social information (preference-weighted N_{with}) and the bet difference.
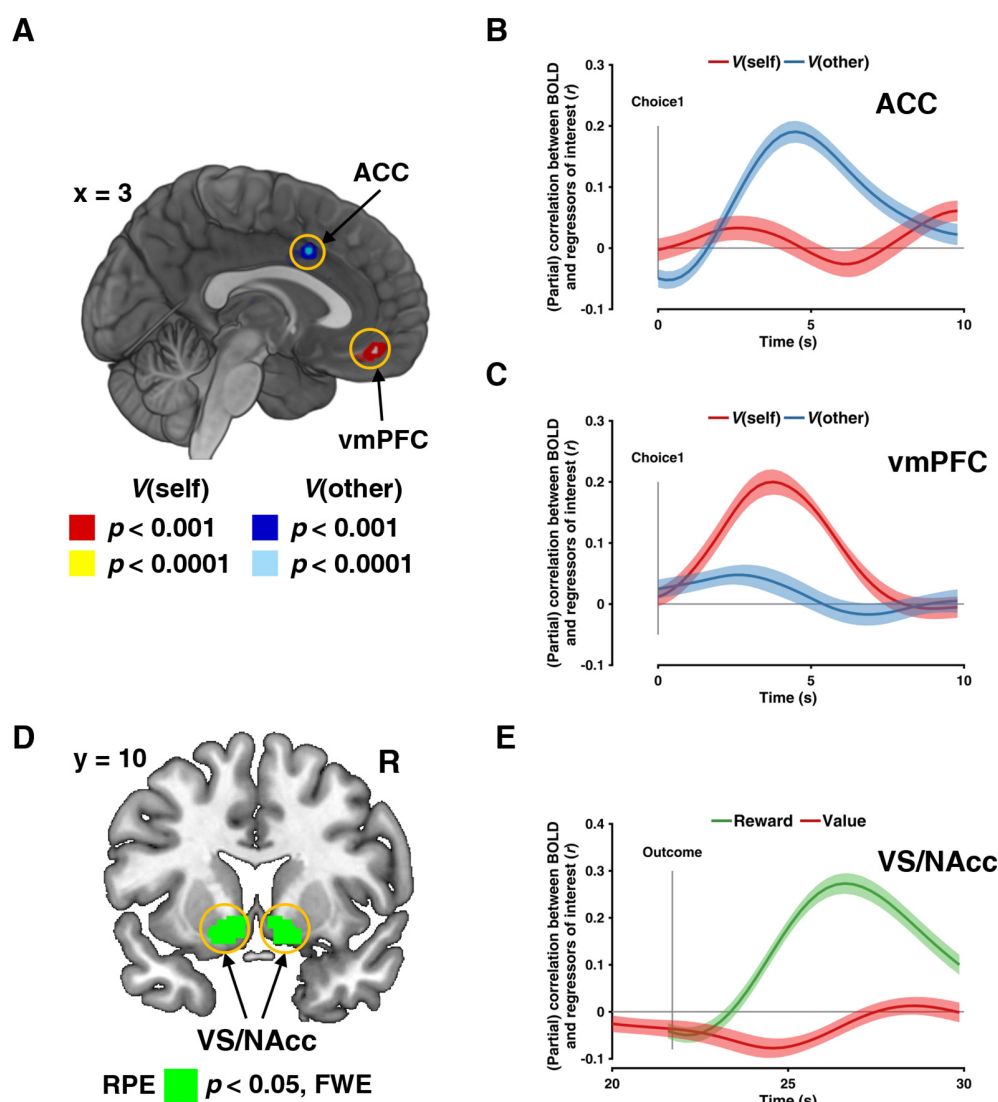
1300

1301

1302

1303 **Figure 3. Neural substrates of dissociable value signals and reward prediction error. (A)** The
1304 neural representation of V_{self} and V_{other} are encoded in the vmPFC (red/yellow) and the ACC
1305 (blue/light blue), respectively. Sagittal slice at x = 3. **(B, C)** Time series estimates (Behrens et al.,
1306 2008) demonstrate a double dissociation of the neural signatures of the value signals. The vmPFC
1307 is positively correlated with V_{self}, but not with V_{other}, whereas the ACC is positively correlated with
1308 V_{other}, but not with V_{self}. **(D)** The neural representation of reward prediction error (RPE) is encoded
1309 in the VS/NAcc. Coronal slice at y = 10. **(E)** The time series in the left VS/NAcc is sensitive to
1310 both component of the RPE: positive correlation with reward (green line), and negative correlation
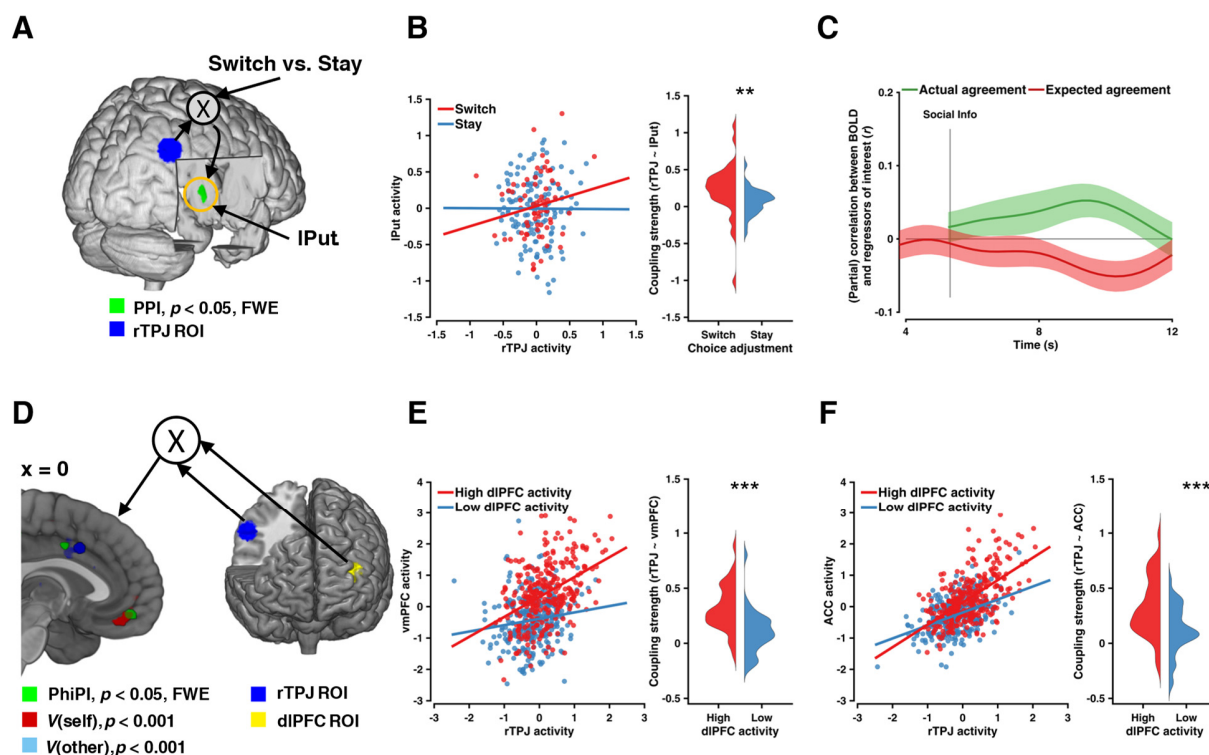1311 with value (red line).

1312

Figure 4. Functional connectivity between reward-related regions and social-related regions. **(A)** The functional connectivity between the left putamen (green) and the seed region rTPJ (blue) is modulated by the choice adjustment (switch vs. stay). **(B)** Correlation of activity in seed and target region for both switch and stay trials in an example subject and histogram of coupling strength across all participants for switch and stay trials. **(C)** The BOLD time series in the left putamen (PPI target) exhibits a social prediction error (positive correlation with the actual agreement, and negative correlation with the expected agreement) (mean effect across participants $\pm$ SEM). **(D)** Two seed regions, the rTPJ (blue), which responds to the social information, and the left dlPFC (yellow), which encodes the choice adjustment, elicit connectivity activations in the vmPFC and the pMFC (both in green), which partially overlap with the latent value signals (i.e., V_{self}; red; and V_{other}; blue), as in Figure 3A. Sagittal slice at x = 0. **(E, F)** Correlation plots of seed and target regions for both high and low dlPFC activity in an example subject and histograms of seed-target coupling strengths across all participants for high and low dlPFC activity.
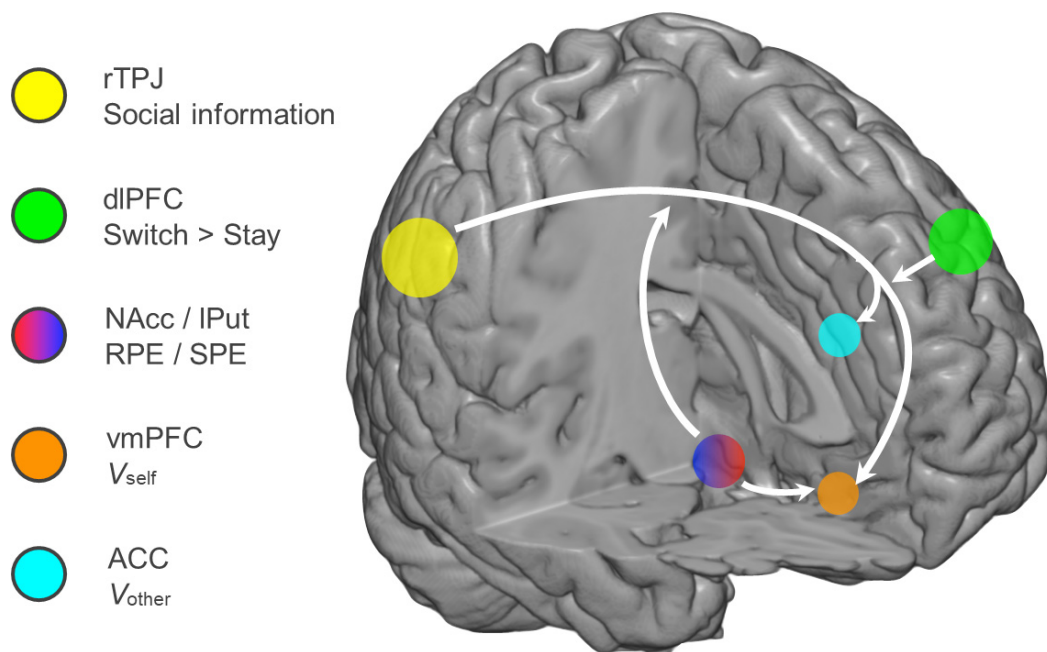
1330

1331

1332

1333 **Figure 5. Schematic of the of the network supporting social influence in decision-making as**
1334 **uncovered in this study** (for details see main text)**.**

1335

1336 **Table 1. Candidate computational models and model comparison**

Class	Model	Description	ΔLOOIC
Non-social models	M1a	simple RL	0
	M1b	M1a + fictitious update	−1245
Social models with instantaneous effect	M2a	M1a + instantaneous social influence	−1893
	M2b	M1b + instantaneous social influence	−2889
Social models with instantaneous effect and observational learning	M3	M2b + OL (others' RL update)	−3089
	M4	M2b + OL (others' action preference)	−2869
	M5	M2b + OL (others' current reward)	−3203
	M6a	M2b + OL (others' cumulative reward)	−3450
	M6b	M2b + OL (others' cumulative reward) + bet1	−3507

1337

1338 Note: RL = reinforcement learning, OL = observational learning. LOOIC = leave-one-out

1339 information criterion; lower LOOIC value indicates better out-of-sample predictive accuracy. M6b

1340 (in bold) is the winning model.
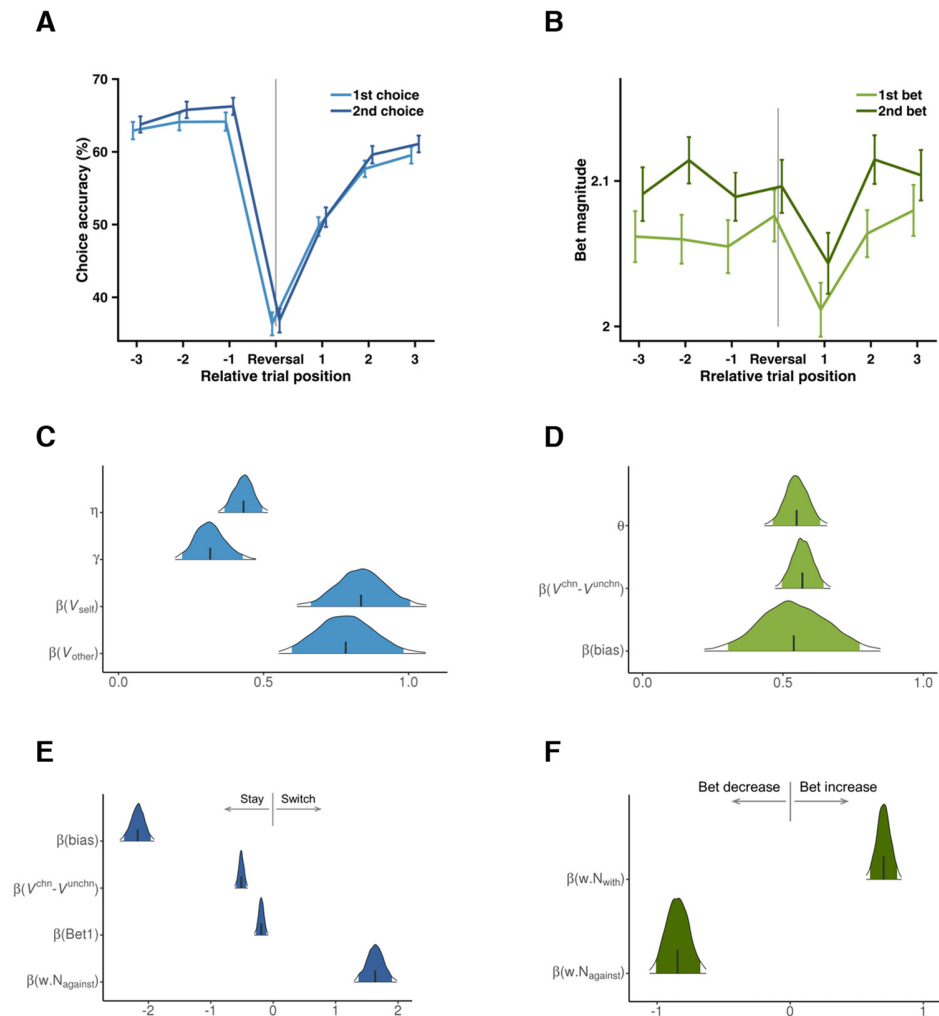
1341

Supplementary Information

Figure S1.

Social information enhances learning and associated model parameters, related to Figures 1-2 and Star Methods. **(A)** Choice accuracy across reversal. Participants' choice accuracy as a function of the choice type (1st choices, light blue; 2nd choices, dark blue) and the trials positions relative to the reversal (−3, −2, −1, 0, 1, 2, 3). All error bars indicate within-subject SEM. The gray vertical line indicates the reversal. **(B)** Bet magnitude across reversal. Participants' bet magnitude as a function of the bet type (1st bet, light green; 2nd bet, dark green) and the trials positions relative to the reversal (−3, −2, −1, 0, 1, 2, 3). All error bars indicate within-subject SEM. The gray vertical line indicates the reversal. **(C-F)** Model parameters. Parameters' posterior density for (C) first choice, (D) second choice, (E) first bet, and (F) second bet. Shaded areas depict 95% of the highest density interval (HDI).
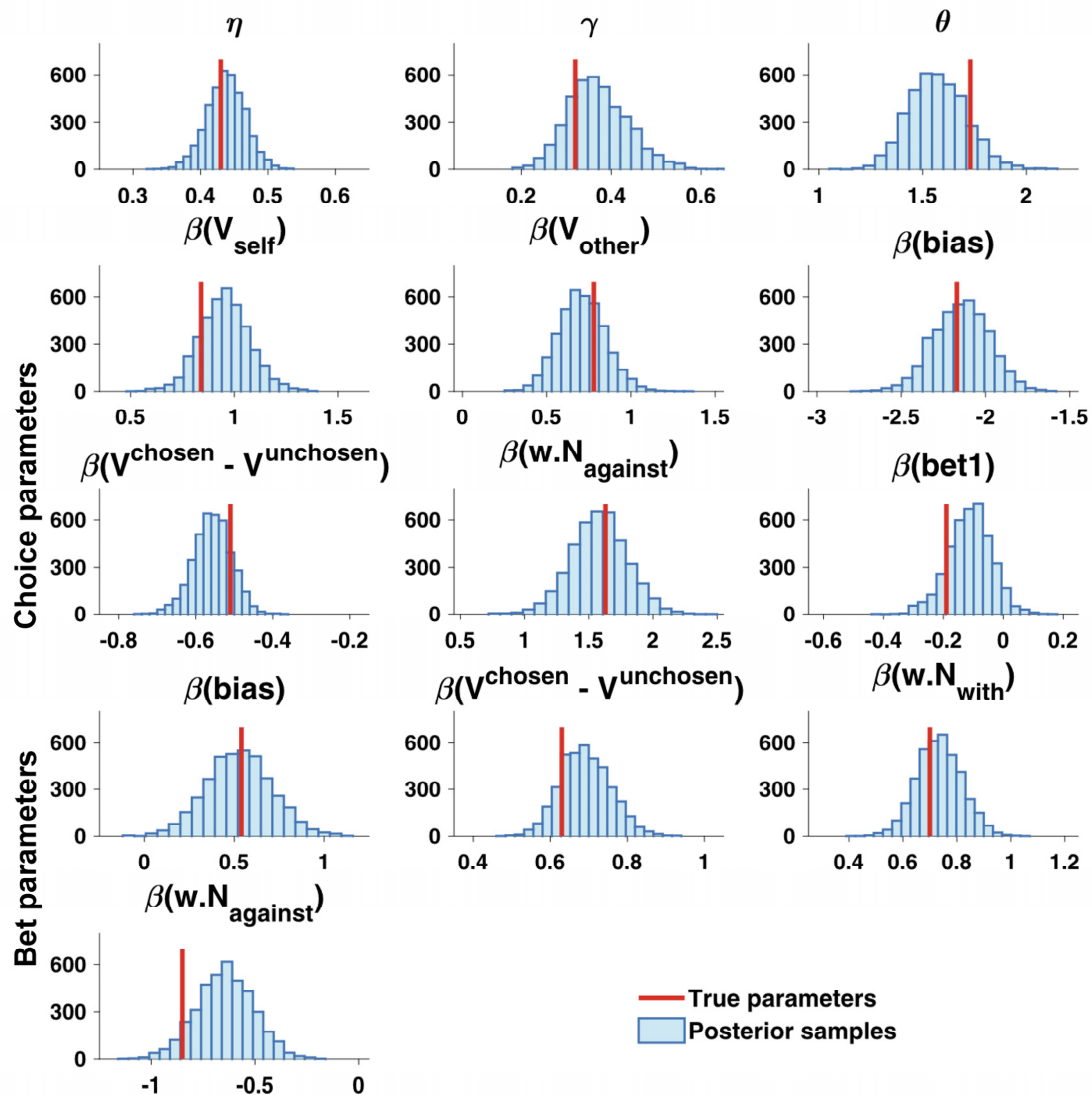
Figure S2.

Parameter recovery, related to Figure2 and Star Methods. A random sample of group-level parameters is drawn from the joint posterior group-level distribution to simulate 80 synthetic participants' behavioral data for the social influence task. All parameters could be accurately and selectively recovered ("true" parameters falling between the 95% HDI of each posterior distribution), suggesting the proper identifiability of model parameters.
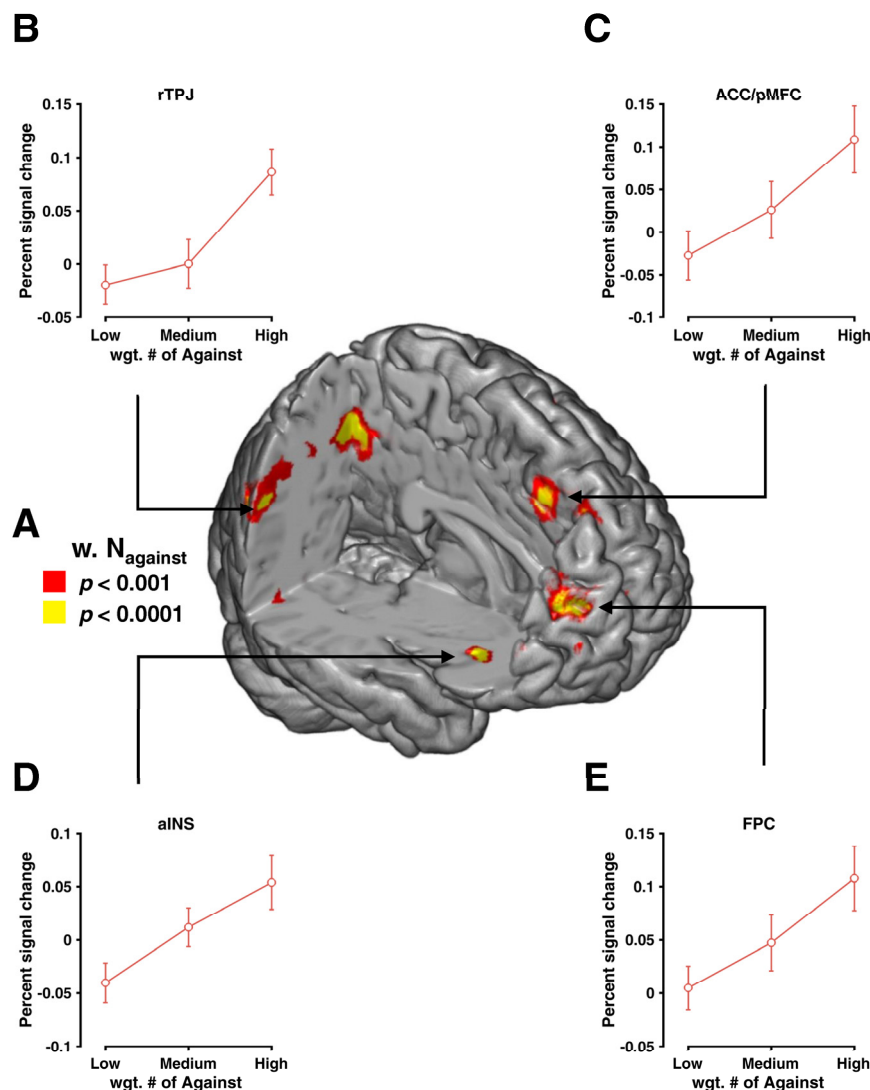
Figure S3.

Neural substrates of dissenting social information, related to Figures 4-5. **(A)** Activities (red/yellow) in the rTPJ, the ACC/pMFC, bilateral aINS and the FPC scaled parametrically with the dissenting social information (i.e., $w.N_{against}$). Color coding indicates the SPM thresholds corresponding to $p < 0.001$ and $p < 0.0001$ uncorrected. **(B-E)** Percent signal change (PSC) of BOLD time series in the corresponding ROI for trials in which w. $N_{against}$ is low, medium, and high (33rd, 66th, and 100th percentile range). All show a significant linear increasing trend (F test, $ps < 0.001$). For the sub-figure D, we chose the right aINS because its supra-threshold cluster was larger than the left one (Table S4). Using the left aINS yielded similar results. All error bars indicate SEM.
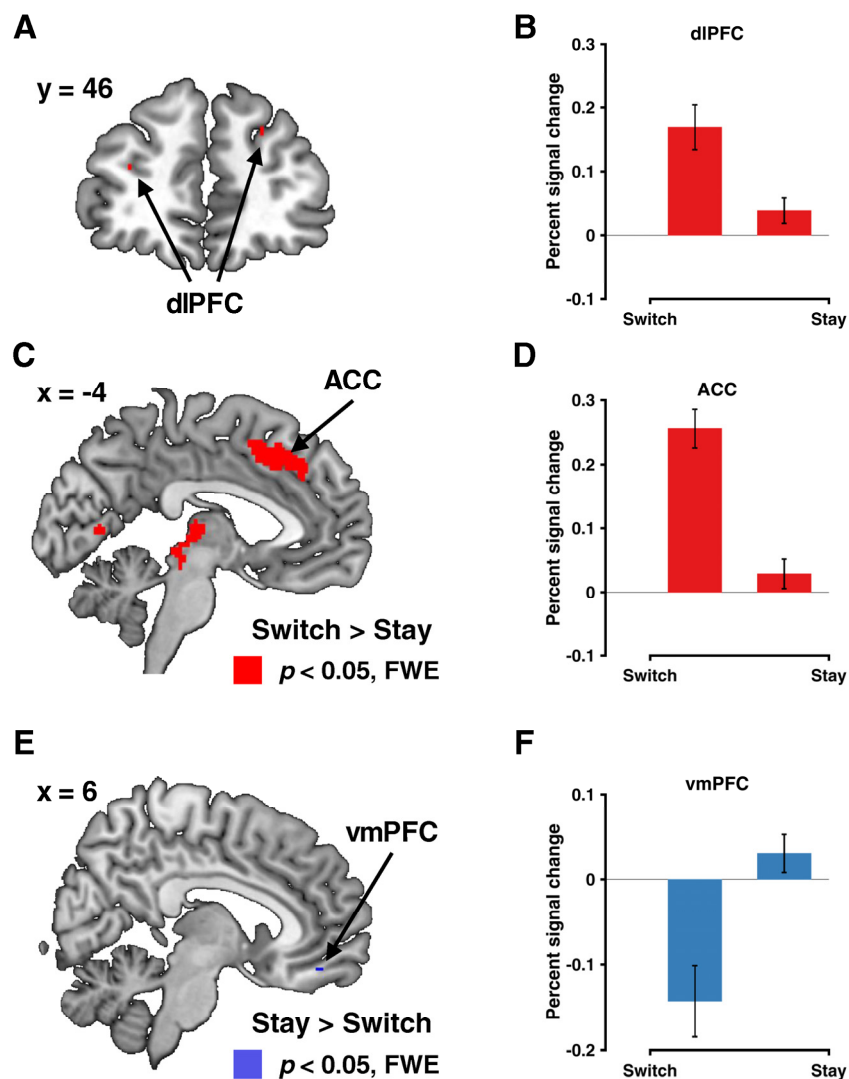
3

Figure S4.

Neural substrates of choice adjustment, related to Figures 4-5. **(A, C)** Bilateral dlPFC and the ACC (red) are more activated when individuals made a switch during the choice adjustment. **(E)** The vmPFC (blue) is more activated when individuals made a stay during the choice adjustment. Color coding indicates the SPM threshold corresponding to $p < 0.05$ FWE corrected. **(B, D, F)** Percent signal change (PSC) of BOLD time series in the corresponding ROI for switching trials as opposed to staying trials. All show a significant difference (paired t-test, ps < 0.001). For the sub-figure B, we chose the left dlPFC because its supra-threshold cluster was larger than the right one (Table S4). Using the right dlPFC yielded similar results. All error bars indicate SEM.
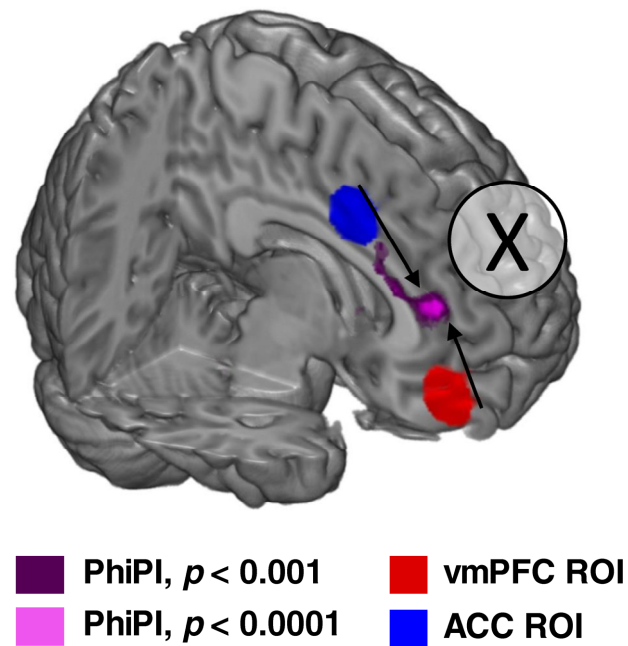
PhiPI, *p* < 0.001   **vmPFC ROI**
PhiPI, *p* < 0.0001   **ACC ROI**

Figure S5.

PhiPI between vmPFC and ACC, related to Figures 3-5. Two seed regions, the vmPFC (red), which is associated with $V_{\text{self}}^{\text{chosen}}$, and the ACC (blue), which is linked with $V_{\text{other}}^{\text{chosen}}$, elicit connectivity activations in the (rostral) mPFC (magenta). Color coding indicates the SPM thresholds corresponding to $p < 0.001$ uncorrected, $p < 0.0001$ uncorrected, respectively.
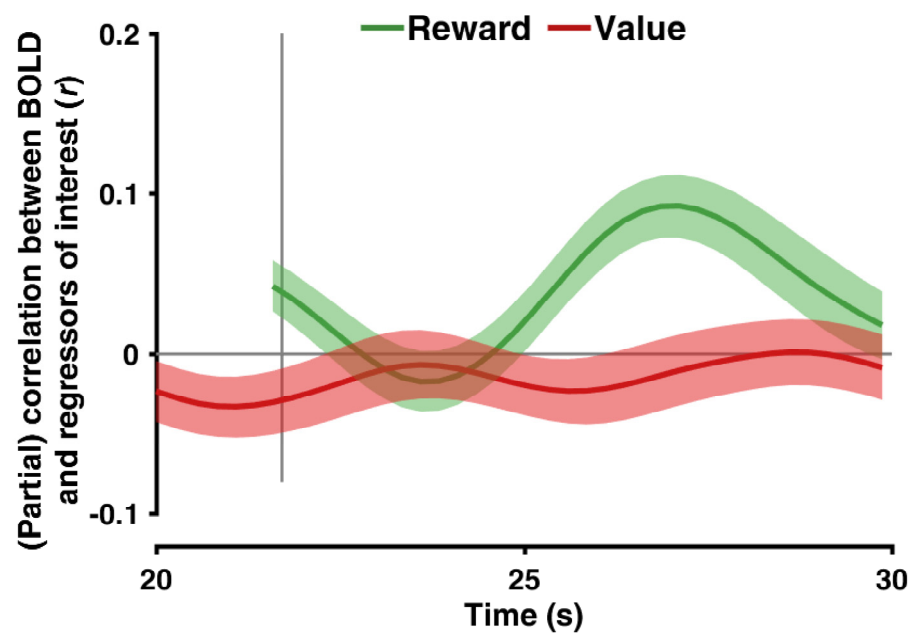
Figure S6.

(Non-)Association between left putamen and components of reward prediction error, related to Figure 4. The BOLD time series in the left putamen is only positively correlated with the actual reward outcome (green line; $p < 0.0001$, permutation test), but not negatively correlated with the expected outcome (i.e., value; red line, $p = 0.4854$, permutation test).

Table S1.

Mixed-effect models for the measurement of choice switch probability and bet difference, related to Figure 1 and Figure S1.

Model	AIC	*p*
Choice Switch Probability		
y ~ dir*con + (1|sub)	−506.75	--
y ~ dir*con + (1|sub) + (1|gender) + (1|group)	−502.84	.956
y ~ dir*con + (1|sub) + (1|gender) + (1|group) + (1|dir:sub) ←	−526.13	< .001
y ~ dir*con + (1|sub) + (1|gender) + (1|group) + (1|con:sub)	−500.84	1.000
y ~ dir*con + (1|sub) + (1|gender) + (1|group) + (1|dir:sub) + (1|con:sub)	−524.13	< .001
Bet Difference		
y ~ dir*con + (1|sub) ←	101.21	--
y ~ dir*con + (1|sub) + (1|gender) + (1|group)	105.21	1.000
y ~ dir*con + (1|sub) + (1|gender) + (1|group) + (1|dir:sub)	107.21	.983
y ~ dir*con + (1|sub) + (1|gender) + (1|group) + (1|con:sub)	107.21	1.000
y ~ dir*con + (1|sub) + (1|gender) + (1|group) + (1|dir:sub) + (1|con:sub)	109.21	.937

Note: Models are specified in the "lme4" syntax: "A*B" denotes the main effect plus the interaction effect between variables A and B; "(1|var)" denotes the single random effect; "(1|A*B)" denotes the interaction random effect between variables A and B. "dir" = direction (with vs. against the group). "con" = group consensus level (2:2, 3:1, 4:0). "sub" = subject. "gender" = gender (female vs. male). "group" = measurement group (fMRI vs. behavioral). AIC = Akaike Information Criterion. Lower AIC value indicates better model. *p* values are calculated with χ^2 test. Models that best capture the random structure of the data are indicated by a left-pointing arrow sign (←).

Table S2.

Mixed-effect models for the measurement of choice accuracy and bet magnitude, related to Figure 1 and Figure S1.

Model	AIC	p
Choice Accuracy		
`y ~ typ*pos + (1|sub)` ←	−1559.9	--
`y ~ typ*pos + (1|sub) + (1|gender) + (1|group)`	−1555.9	1.000
`y ~ typ*pos + (1|sub) + (1|gender) + (1|group) + (1|typ:sub)`	−1553.9	1.000
Bet Magnitude		
`y ~ typ*pos + (1|sub)`	713.16	--
`y ~ typ*pos + (1|sub) + (1|gender) + (1|group)`	712.83	.115
`y ~ typ*pos + (1|sub) + (1|gender) + (1|group) + (1|typ:sub)` ←	656.75	< .001

Note: Models are specified in the "lme4" syntax: "A*B" denotes the main effect plus the interaction effect between variables A and B; "(1|var)" denotes the single random effect; "(1|A*B)" denotes the interaction random effect between variables A and B. "typ" = choice type (1st choice vs. 2nd choice). "pos" = relative trial position (−3, −2, −1, reversal, +1, +2, +3). "sub" = subject. "gender" = gender (female vs. male). "group" = measurement group (fMRI vs. behavioral). AIC = Akaike Information Criterion. Lower AIC value indicates better model. p values are calculated with χ^2 test. Models that best capture the random structure of the data are indicated by a left-pointing arrow sign (←).

Table S3.

Neural substrates of value and reward prediction error (RPE) signals, related to Figure 3.

Contrast	Region	MNI coordinates (peak)			Cluster size	Z_{max}
		x	y	z		
V_{self}^{chosen}	vmPFC (BA11)	4	46	−14	49[a]	3.91[*]
V_{other}^{chosen}	ACC (BA32)	2	10	36	55[a]	3.94[*]
RPE	left VS/NAcc (BA48)	−10	8	−10	199[b]	7.07[**]
	right VS/NAcc (BA52)	12	10	−12	171[b]	7.35[**]
	vmPFC (BA10)	−10	62	2	62[b]	6.01[**]

Note: [*]: TFCE with small volume correction (SVC), at $p < 0.05$, FWE corrected; [**]: whole-brain TFCE correction, at $p < 0.05$, FWE corrected; [a]: cluster size obtained at $p < 0.001$, uncorrected; [b]: cluster size obtained at $p < 0.05$, FWE corrected. V_{self}^{chosen} = chosen "self value" updated from individuals' own valuation; V_{other}^{chosen} = chosen "other value" updated from the others' cumulative reward history; RPE = reward prediction error; vmPFC = ventromedial prefrontal cortex; ACC = anterior cingulate cortex; VS = ventral striatum; NAcc = nucleus accumbens.

Table S4.

Neural substrates of instantaneous social information and behavioral adjustment, related to Figure 4, Figure S1 and Figure S4.

Contrast	Region	MNI coordinates (peak)			Cluster size	Z_{max}
		x	y	z		
w. $N_{against}$	rTPJ (BA39)	50	−60	34	214[a]	4.44[**]
	lTPJ (BA39)	-48	-62	30	167[a]	3.06[**]
	ACC/pMFC (BA8)	4	28	44	238[a]	5.03[**]
	left aINS (BA13)	−30	18	−14	56[a]	3.90[**]
	right aINS (BA13/47)	32	24	−10	163[a]	5.13[**]
	FPC (BA10)	22	60	18	140[a]	4.97[**]
	Frontal-mid L (BA10)	−26	50	16	124[a]	4.75[**]
	right-Fusiform (BA37)	30	−68	−12	238[a]	5.44[**]
SwSt	left dlPFC (BA10)	−32	48	16	27[b]	5.23[**]
	right dlPFC (BA9)	26	42	32	21[b]	5.56[**]
	ACC (BA8)	−4	16	44	166[b]	6.13[**]
	left Thalamus (BA50)	−12	−18	10	156[b]	6.50[**]
	left Lingual (BA19)	−24	−68	−10	113[b]	6.81[**]
	left su. Occip. (BA19)	28	−78	20	110[b]	6.87[**]
	left su. Pariat. (BA7)	−26	−48	50	117[b]	6.39[**]
StSw	vmPFC (BA11)	6	46	−16	4[b]	5.07[**]
	left mid. Tem. (BA22)	−62	−28	6	7[b]	5.68[**]
	right rol. Oper. (BA6)	58	2	8	8[b]	5.28[**]

Note: [**]: whole-brain TFCE correction, at $p < 0.05$, FWE corrected; [a]: cluster size obtained at $p < 0.001$, uncorrected. [b]: cluster size obtained at $p < 0.05$, FWE corrected. w. $N_{against}$ = preference-weighted number of against options from the other co-players; SwSt = switch > stay; StSw = stay > switch. rTPJ = right Temporal Parietal Junction; ACC = anterior cingulate cortex; pMFC = posterior medial frontal cortex; aINS = anterior insula; FPC = frontopolar cortex; dlPFC = dorsolateral prefrontal cortex; ACC = anterior cingulate cortex; vmPFC = ventromedial prefrontal cortex; su. Occip. = superior occipital gyrus. su. Pariat. = superior parietal lobule; mid. Tem. = middle temporal gyrus; rol. Oper. = Rolandic Operculum.

Table S5.

Functional connectivity, related to Figure 4.

Connectivity	Region	MNI coordinates (peak)			Cluster size	Z_{max}
		x	y	z		
vmPFC ~ ACC	mPFC (BA32)	10	40	10	170[a]	4.62**
	l-Caudate (BA48)	−10	4	20	130[a]	4.87**
	r-Insula (BA13)	38	6	4	191[a]	5.18**
rTPJ ~ SwSt	l-putamen (BA49)	−20	12	−4	104[b]	6.08**
	l-su.Pra. (BA40)	−56	−34	36	37[b]	6.00**
	l-Thalam. (BA50)	−6	−14	10	26[b]	5.80**
rTPJ ~ left dlPFC	vmPFC (BA10)	0	48	−12	23[b]	5.26**
	ACC (BA24)	0	0	40	12[b]	5.12**
	r-Insula (BA13)	44	6	−10	214[b]	6.57**
	l-Insula (BA13)	−46	8	−8	185[b]	6.37**

Note: **: whole-brain TFCE correction, at $p < 0.05$, FWE corrected; [a]: cluster size obtained at $p < 0.001$, uncorrected; [b]: cluster size obtained at $p < 0.05$, FWE corrected. SwSt = switch > stay; vmPFC = ventromedial prefrontal cortex; ACC = anterior cingulate cortex; rTPJ = right temporal parietal junction; dlPFC = dorsolateral prefrontal cortex; mPFC = medial prefrontal cortex; l-putamen = left putamen; ACC = anterior cingulate cortex; Su.Pra. = supramarginal gyrus; Thalam. = Thalamus.