Behavioral/Cognitive

Computational Substrates of Social Value in Interpersonal Collaboration

Dominic S. Fareri,[1]* Luke J. Chang,[2]* and Mauricio R. Delgado[3]

[1]Gordon F. Derner Institute of Advanced Psychological Studies, Adelphi University, Garden City, New York 11530, [2]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire 03755, and [3]Department of Psychology, Rutgers University, Newark, New Jersey 07102

Decisions to engage in collaborative interactions require enduring considerable risk, yet provide the foundation for building and maintaining relationships. Here, we investigate the mechanisms underlying this process and test a computational model of social value to predict collaborative decision making. Twenty-six participants played an iterated trust game and chose to invest more frequently with their friends compared with a confederate or computer despite equal reinforcement rates. This behavior was predicted by our model, which posits that people receive a social value reward signal from reciprocation of collaborative decisions conditional on the closeness of the relationship. This social value signal was associated with increased activity in the ventral striatum and medial prefrontal cortex, which significantly predicted the reward parameters from the social value model. Therefore, we demonstrate that the computation of social value drives collaborative behavior in repeated interactions and provide a mechanistic account of reward circuit function instantiating this process.

Key words: **collaboration; medial prefrontal cortex; social network; social value; trust; ventral striatum**

Introduction

Collaboration is essential to our social life, providing the foundation for advancing our economic, technological, political, and personal landscapes. One critical aspect of collaboration is the construct of trust, which can be described as assuming mutual risk with a relationship partner to attain an interdependent goal (Simpson, 2007). However, collaborations are not just about attaining self-interested goals, but also about fostering interpersonal relationships. Relationships are intrinsically rewarding and help to fulfill a basic social need to belong (Baumeister and Leary, 1995). Maintaining stable and close relationships promotes positive physical and mental health outcomes (Uchino, 2009), highlighting the importance of understanding the mechanisms facilitating collaborations and sustaining relationships.

Modern behavioral economic models have made dramatic improvements to classical economic theory in predicting collaborative behavior by incorporating social preferences such as considering other's intentions (Rabin, 1993) or payoff outcomes (Fehr and Schmidt, 1999). In addition, prior expectations can bias our willingness to take collaborative risks such that we overweigh information consistent with expectations when deciding to trust someone (Delgado et al., 2005; van't Wout and Sanfey, 2008; Fareri et al., 2012a) in accordance with confirmation bias (Doll et al., 2009). Importantly, these expectations appear to be malleable and are updated after receiving feedback using a prediction error computation (Chang et al., 2010) processed in the ventral striatum (Fareri et al., 2012a).

It remains unclear, however, how collaborations might be influenced by strong priors based on years of repeated interactions. One possibility is that continued collaboration with close others may be driven by a strong prior expectation of reciprocation, which would be associated with less prediction error and ventral striatal activation after reciprocation (Fouragnan et al., 2013). Alternatively, reciprocation may be desirable in close relationships (Rilling et al., 2002; Phan et al., 2010) and may evoke a larger reward signal as it strengthens an existing social bond. This would imply that the social value of reciprocation might be modulated by aspects of the relationship; for example, degree of closeness or perceived trustworthiness.

We formalized and tested these competing hypotheses using a computational modeling approach within the context of a repeated trust game (Delgado et al., 2005) in which we manipulated the social network status of participants' partners (Fareri et al., 2012b). Participants made investment decisions with a close friend (in-network), a confederate (out-of-network), and a computer (nonsocial control) while undergoing fMRI. We formalized an expectation-learning model using a standard reinforcement-learning (RL) model with strong prior expectations to test whether participants' decisions to invest with a close friend were primarily motivated by a strong prior trustworthiness belief. We compared this with a new social value model based on expected value theory in which the value term is composed of a linear combination of self-interested (i.e., financial) and social value

Received Nov. 21, 2014; revised April 10, 2015; accepted April 16, 2015.

Author contributions: D.S.F. and M.R.D. designed research; D.S.F. performed research; D.S.F. and L.J.C. analyzed data; D.S.F., L.J.C., and M.R.D. wrote the paper.

This work was supported by the National Institute of Mental Health–National Institutes of Health (Grant R01MH084081 to M.R.D.).

The authors declare no competing financial interests.

*D.S.F. and L.J.C. contributed equally to this work.

Correspondence should be addressed to Dr. Mauricio R. Delgado, Rutgers University, Department of Psychology, Smith Hall, Rm. 340, 101 Warren Street, Newark, NJ 07102. E-mail: delgado@psychology.rutgers.edu.

DOI:10.1523/JNEUROSCI.4775-14.2015

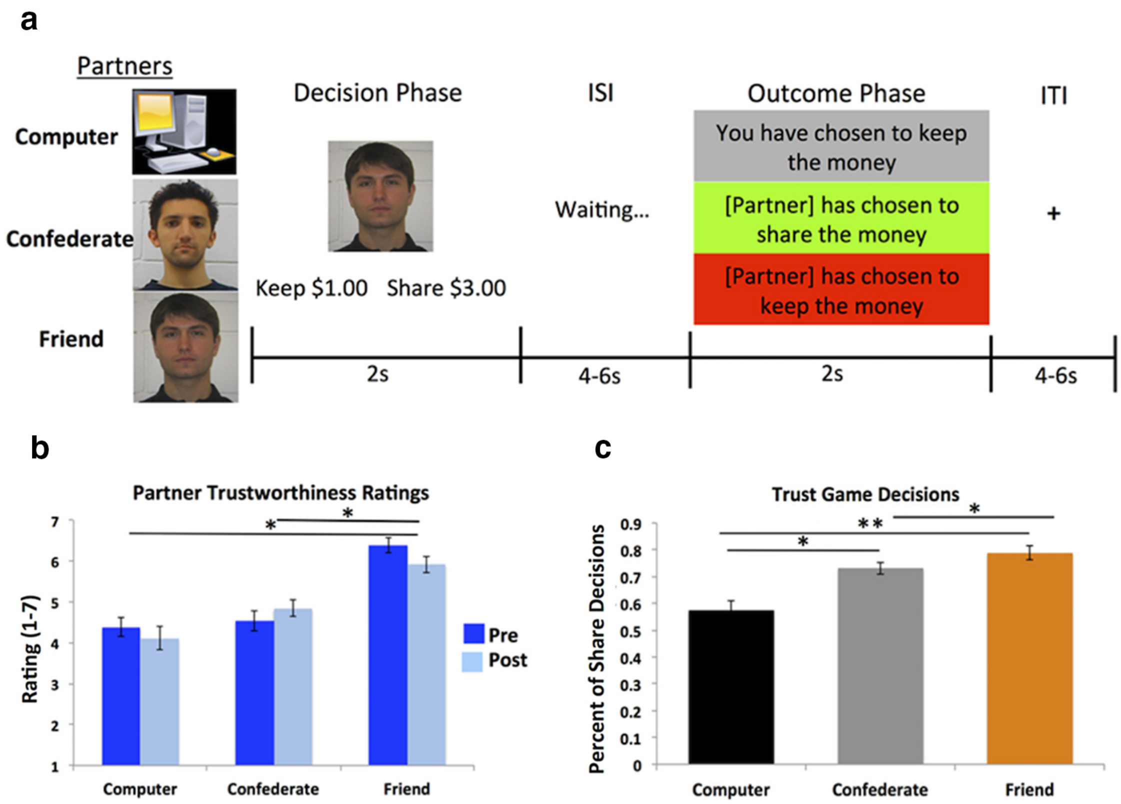

Figure 1. Task schematic, manipulation check, and trust decisions. ***a***, MRI participants played a trust game with three different partners: a close friend (in-network), a confederate (out-of-network), and a computer (nonsocial control). Participants were endowed with $1.00 on each trial and chose whether to keep the money for themselves, leaving the partner with $0, or to share/invest with that partner. Decisions to share resulted in the partner receiving a tripled amount of money ($3.00). After submitting their decision, a screen that said "waiting" appeared, during which participants believed that they were waiting for their partner's decision. Partners' decisions to share resulted in an even split of the $3.00 investment, whereas decisions to keep resulted in $0 being returned to the participant. ***b***, Participants ($n = 26$) assessed each partner's trustworthiness before (pre) and after (post) the task "How trustworthy is this partner?" ***c***, Percentage of trials in which participants shared on average across the experiment conditional on the partner context. Participants shared significantly more with their close friends compared with the confederate and computer and more often with the confederate than the computer. $^{**}p < 0.0001$; $^{*}p < 0.005$ (±SEM).

(i.e., trustworthiness) signals and the probability term dynamically updates with beliefs about the likelihood of partner reciprocation. This social value model tests whether participants' decisions to invest with a close friend are primarily driven by a social reward bonus feedback signal based on the subjective quality of the relationship as opposed to strong prior expectations. We hypothesized the following: (1) that the social value model would predict collaborative behavior better than the standard expectation model and (2) that we would see a corresponding social value signal in neural circuits of reward [e.g., striatum, medial prefrontal cortex (mPFC)].

Materials and Methods

Participants

Twenty-nine sex-matched participant pairs (16 female) from Rutgers–Newark and the surrounding area took part in this study. Three participant pairs were excluded from analysis due to excessive head motion/image artifact, failure to attend to task due to sleeping, or never experiencing one of the conditions during the task. Analyses were conducted on the remaining 26 MRI participants (14 female; mean age = 21.36, SD = 3.67). All participants provided informed consent before taking part in the experiment and all were screened for history of psychiatric illness and head trauma. This Institutional Review Board of Rutgers University approved this study.

Experimental paradigm

We applied a social network manipulation (Fareri et al., 2012b; Fareri and Delgado, 2014a) to an iterated economic trust game (Fig. 1*a*). MRI participants interacted in this game with a same-sex close friend whom they brought to the experimental session (in-network), a sex-matched confederate (out-of-network), and a computer (nonsocial control). Because we expected MRI participants to feel close to their friend, we assessed social closeness via a simple measure consisting of pairs of overlapping circles, one labeled self and one labeled other using the Inclusion of Other in Self Scale (IOS; Aron et al., 1992). Increased overlap suggests increased closeness. MRI participants chose the pair of circles that best characterized their relationship with their friend. Participant pairs were then brought to the Rutgers University Brain Imaging Center (RUBIC, Newark, NJ) and introduced to a sex-matched confederate who was portrayed as an additional participant. In reality, the confederate was a laboratory member whose identity was concealed until the end of the session. Before the start of the scan session, we asked MRI participants to make subjective ratings of trustworthiness for each partner using a 7-point Likert scale where 1 = not at all and 7 = a lot. While this was being completed, a facial photograph was taken of the same-sex friend and programmed into the task as a stimulus. We also asked them to fill out the IOS with respect to the confederate and the computer.

The MRI participant, close friend, and confederate were subsequently seated together in the control room and told that they would be playing the investment game (i.e., an iterated trust game; Delgado et al., 2005; Fareri et al., 2012a). The MRI participant was designated the investor and told that s/he would play the game with one partner on each trial (Fig. 1*a*). MRI participants were endowed with $1.00 on each trial, which they could keep, signaling the end of the trial, or share with their partner. A choice to share was described as an investment, resulting in a tripling of the money to $3.00 for the partner on a given trial (Berg et al., 1995; Delgado et al., 2005); the respective partner could decide to keep all $3.00 or share it back evenly with the MRI participant ($1.50 each). Both MRI participants and their human partners underwent a series of practice trials to ensure understanding of the task. Once the MRI participant was situated in the scanner, however, the friend and confederate were instructed that they did not actually have to take part in the task: their responses were preprogrammed to demonstrate equivalent reputation during the task (see below).

Trials consisted of a decision and outcome phase (Fig. 1*a*). During the decision phase (2 s), a photo of 1 of the 3 partners was presented on the screen. MRI participants chose to keep or share via button presses on an MRI-compatible fiber optic response pad (Current Designs). A jittered interstimulus interval (ISI; 4–6 s) followed, during which the word "waiting" was presented on the screen; MRI participants believed that their decision was being transmitted to the computer in the control room at which their partners were seated so that they would get the opportunity to respond if the money was shared. Partner decisions were revealed during the outcome phase (2 s) and all trials were separated by a jittered intertrial interval (6–8 s). Missed trials (no response in the decision phase) were indicated by a "#" symbol after the ISI. MRI participants were compensated for their participation at a rate of $25/experimental hour plus bonus payment based on realized outcomes of trials from two randomly chosen task runs. Any missed trials that were included in these runs were not eligible for bonus payment.

The task consisted of 72 trials in total, evenly distributed across six functional runs. Trial order was counterbalanced across participants. Twenty-four trials per partner condition were randomly administered across these runs. Partner responses were preprogrammed: the reinforcement schedule was set to be equivalent so that all partners reciprocated on 50% of trials in which MRI participants chose to invest. This allowed for an equivalent proportion of positive (i.e., reciprocation) and negative (i.e., defection) outcomes to isolate the effects of the social partner on neural representations of outcome value. MRI participants assessed partner trustworthiness after the task on a 7-point Likert scale (1 = not at all, 7 = a lot). Participants were then debriefed and compensated. MRI participants were paid as described above; their friends were paid at a rate of $10/h.

Behavioral analysis

Subjective ratings. The IOS ratings and pretask/posttask trustworthiness ratings for each partner served as a social network manipulation check. IOS ratings were entered into a one-way repeated-measures ANOVA. Pretask and posttask trustworthiness ratings were entered into a 2 (time: pre/post) × 3 (partner) repeated-measures ANOVA. A Greenhouse–Geiser correction was applied to tests violating conditions of sphericity. *Post hoc* comparisons were conducted and corrected for multiple comparisons via the sequential Bonferroni method (Holm, 1979; Rice, 1989).

Trust decisions. We examined MRI participants' decisions to keep or share as a function of partner using a mixed-effects logistic regression with randomly varying slopes and intercepts. We also probed differences in log-transformed reaction times as a function of decision and partner via a mixed-effects linear regression with randomly varying intercepts. These analyses were conducted using the lmerTest (Kuznetsova et al., 2014) and LME4 (Bates et al., 2014) packages in the R statistical language. *Post hoc* comparisons were conducted using the sequential Bonferroni method (Holm, 1979; Rice, 1989).

Computational models. We used computational models to test specific cognitive mechanisms that might facilitate observed behavioral effects. This analytic approach has been applied successfully to investigations of social learning (Behrens et al., 2008; Chang et al., 2010; Jones et al., 2011; Kishida and Montague, 2012) and we have demonstrated previously that people appear to use prediction error learning (Schultz et al., 1997; Sutton and Barto, 1998) to update dynamically their beliefs about a relationship partner's trustworthiness during repeated interactions (Chang et al., 2010; Fareri et al., 2012a). We formalized and tested three main computational models: (1) a baseline expected value model with no learning components, (2) a RL model with prior expectations, and (3) a social value model of collaborative behavior. We also tested three additional control models to examine potential alternative interpretations of our data: (1) a partner reciprocation value model, (2) a loss–gain RL model (Fareri et al., 2012a), and (3) a loss–gain RL partners model (Fareri et al., 2012a).

Main models

Expected value model. We used a decision theory framework to formalize a baseline model, which posits that participants make their decisions by maximizing their expected value. The expected value (EV) for an investment (i.e., share) decision on trial t for a given partner context c (e.g., in-network, out-of-network, computer) was represented as the value received from a partner reciprocating scaled by the likelihood of the event occurring (P_c) as follows:

$$EV_c(t) = P_c(t) * (1.5) \tag{1}$$

For the baseline expected value model, we fixed $P_c(t)$ to be 0.5, which reflected the actual probability of reinforcement despite this being unknown to the participants. The expected value (EV_c) was then placed into a softmax function to calculate the probability of a participant investing with a given partner (IP_c) as follows:

$$IP_c(t) = \frac{e^{\frac{EV_c(t)}{\beta}}}{e^{\frac{EV_c(t)}{\beta}} + e^{\frac{1}{\beta}}} \tag{2}$$

Where β varies between 0 and 1 and reflects whether a participant is more likely to behave in a more explorative (e.g., varying choices) or exploitative (e.g., attempting to act in the most advantageous manner, not sampling alternative options). The probability of a participant not investing is equivalent to $1 - IP_c$.

Expectation-learning model. We formalized the expectation-learning model using Equation 1. However, because participants were unaware that the likelihood of partner reciprocation was fixed at 50%, we used a Rescorla–Wagner prediction error (delta) rule (Sutton and Barto, 1998) to update participants' expectations P_c after the partner's behavior γ_c for a given trial t, where $\gamma_c = 1$ when the partner shares and $\gamma_c = 0$ when the partner keeps. The update rule was formalized as follows:

$$P_c(t + 1) = P_c(t) + \alpha * (\gamma_c(t) - P_c(t)) \tag{3}$$

Where α is a free parameter indicating a participant's learning rate, which is bounded between 0 and 1. To account for differing prior expectations based on the established relationship with an in-network partner, we initialized participants' expectations of partner reciprocation (P_c) using their initial trustworthiness ratings for each partner, normalized by the maximum possible trustworthiness rating (7), and scaled by a free parameter ϕ, where $0 < \phi < 5$, as follows:

$$P_c(t = 1) = \min\left(\Phi * \frac{T_c}{\max(T_c)}, 1\right) \tag{4}$$

This formulation provides a strong test of the prior expectations based on idiographic beliefs about the likelihood of each partner reciprocating (Chang et al., 2010).

Social value model. We predicted that, in collaborative interactions, people may receive a social reward bonus after reciprocation that is independent of any monetary outcomes. Therefore, in this model, we extend the standard expected value calculation by adding an additional "social value" term to the value function. Social value was simply represented as the normalized initial perceived trustworthiness of the partner (e.g., pretask trust ratings) $T_c/max(rating)$, scaled by a free parameter θ, where $0 < \theta < 5$, as follows:

$$EV_c(t) = P_c(t) * \left(1.5 + \left(\theta * \frac{T_c}{\max(T_c)}\right)\right) \tag{5}$$

Similar to the expectation-learning model described above, the social value model also allows the probability term P_c to update dynamically as evidence accumulates about the likelihood of a partner reciprocating (Eq. 3). The expected probabilities for each partner P_c were initialized at 0.5 to allow for maximal uncertainty. The expected value was then placed into a softmax function to calculate the probability of a participant investing with a given partner (Eq. 2) and model parameters were estimated for each participant. Therefore, this model differs critically from the expectation model in that collaborative behavior is not driven by prior expectations, but instead by a modified value function in which participants receive a social value bonus at the time of reciprocation.

Control models

Partner reciprocation value model. Our social value model postulates that participants will experience a social reward signal upon reciprocation that is proportional to the subjective quality of the relationship. However, other social preference models make differing predictions regarding possible motivations for collaborative behavior, such as concerns for social efficiency (Hsu et al., 2008) and concerns for others' payoffs (Charness and Rabin, 2002). For example, if people have a general preference for the welfare of a partner, they might opt to make the decision that is associated with the highest total monetary efficiency for all parties regardless of the relationship with the partner. However, our social value model predicts that concern for another's welfare might be modulated by the type of relationship (e.g., in-network vs out-network). To address this, we formulated a model similar to our social value model such that the expected value of sharing with a partner was equivalent to the expected probability of a partner reciprocating, the received financial outcome, and a concern for other value (θ), which we estimated separately for each partner as follows:

$$EV_c(t) = P_c(t) * (1.5 + (\theta_c)) \tag{6}$$

This partner reciprocation value model differs from the original social value model in that: (1) it does not take into account the trustworthiness ratings as part of the bonus calculation and (2) theta can vary across partners. If the estimated theta parameters are greater than zero but similar across partners, then this would provide evidence supporting the general other-regarding preference hypothesis. However, if the theta values vary as a function of a partner such that $\theta_{\text{in-network}} > \theta_{\text{out-network}} > \theta_{\text{computer}}$, then this would provide further support for the social value interpretation.

Loss–gain RL model. We also tested our previously reported three-parameter RL model, which updates beliefs about the likelihood of partner reciprocation separately in the context of gains and losses (for details, see Fareri et al., 2012a). We have demonstrated previously that, in the context of fictional partners whom participants gained initial prior experience with via direct social experience, participants rely more strongly on positive (reciprocation) compared with negative (defection) outcomes during a repeated trust game to update beliefs about their partners. Therefore, we tested whether participants' behavior would be captured in this manner regardless of prior expectations or a social value bonus signal.

Loss–gain RL partners model. Finally, as a supplemental analysis, we tested a computational model to explore whether participants were discounting negative outcomes (e.g., partner defection) experienced within the context of a close friend differently compared with other partners (confederate, computer) (Fareri et al., 2012a). We applied an adapted version of the three-parameter RL model reported previously, which explicitly modeled separate learning rates for positive and negative outcomes (Eq. 6) and separate softmax temperature parameters (Eq. 7) for each partner context as follows:

$$P_c(t+1) = P_c(t) + \alpha_{gain_c} * \max(\gamma - P_c(t), 0) + \alpha_{loss_c} * \min(\gamma - P_c(t), 0) \tag{7}$$

$$IP_c(t) = \frac{e^{\frac{EV_c(t)}{\beta_c}}}{e^{\frac{EV_c(t)}{\beta_c}} + e^{\frac{1}{\beta_c}}} \tag{8}$$

Parameter estimation

Parameters for all models were estimated in MATLAB using the fmincon optimization function separately for each participant by maximizing the log-likelihood of the observed data under the model on a trial-by-trial basis. We reduced the likelihood of the model converging on a local minimum using the rmsearch function and selecting 100 random start locations. Log-likelihood estimates were calculated for each participant by maximizing the following function:

$$LLE = \sum_{t=1}^{n} \log(IP_{c,j}(t)) \tag{9}$$

Where *c* represents the partner, *j* represents a participant's decision to invest or keep, *t* represents the trial, and *n* is the total number of trials. Software for performing model estimation is freely available at http://cosanlab.com/resources.

Model comparisons

Model fits for all models were calculated using the Akaike Information Criterion (Akaike, 1974), which applies a penalty for increased number of free parameters, thus rewarding more parsimonious models. Model fits were compared using a nonparametric Wilcoxon signed-rank test due to deviations from normality as a consequence of noisy estimations. Differences in model-derived outcome bonus parameters in the social value model, the partner reciprocation value model, and the loss–gain RL partners models as a function of partner were examined using repeated-measures ANOVA. *Post hoc t* tests were conducted to probe resulting significant effects. We additionally calculated a measure of percent variance explained via a pseudo R^2 (ρ^2) measure modeled after Camerer and Ho (1999). We calculated a random choice model to fit participant data and used that as a comparison model for the pseudo R^2 calculation as follows:

$$\rho^2 = \frac{\text{AIC}_{\text{random choice}} - \text{AIC}_{\text{model}}}{\text{AIC}_{\text{random choice}}} \tag{10}$$

Parameter recovery

An additional method to evaluate model performance is to calculate how well the estimated model parameters can be recovered using simulations. Importantly, this allows us to assess whether we have a sufficient amount of data to estimate the model parameters reliably. To perform parameter recovery, we simulated data for all models (except the loss–gain RL partners model, which served as a supplemental analysis) for each participant 50 times using the model formulations and original parameters estimated from the behavioral data. For the expectation-learning and social value models, we also used each participant's initial trustworthiness ratings. Decisions to share were determined if the softmax probability to share exceeded $p = 0.5$. For each iteration of the simulation, we refit the model using 10 random start locations to minimize the possibility of the algorithm getting stuck in a local minimum. We then assessed the degree to which the parameters could be recovered by calculating the similarity between the parameters estimated from the behavioral data and the parameters estimated from the simulated data using a Pearson correlation. We report the means and SDs of the similarity (r) for all models except the loss–gain RL partners model across the 50 simulations (Table 1).

fMRI acquisition and analysis. Images were acquired at RUBIC on a 3T Siemens Magnetom Trio whole-body scanner. Anatomical images were collected with a T1-weighted MPRAGE sequence (256 × 256 matrix; FOV = 256 mm; 176 1 mm sagittal slices). Functional images were acquired with a single shot gradient EPI sequence (TR = 2000 ms, TE = 30 ms, FOV = 192, flip angle = 90°, bandwidth = 2232 Hz/Px, echo spacing = 0.51) comprising 33 contiguous oblique-axial slices (3 × 3 × 3 mm voxels) parallel to the anterior–posterior commissure line. Data were preprocessed and analyzed with BrainVoyager QX version 2.6 (Brain Innovation). Standard preprocessing steps were applied: 3D motion correction (six parameters), slice-scan time correction (cubic spline interpolation), 3D Gaussian spatial smoothing (4 mm FWHM), voxelwise linear detrending, and temporal high-pass filtering of frequencies (3 cycles per time course). Structural and functional data were transformed to standard Talairach stereotaxic coordinate space (Talairach and Tournoux, 1988).

Our primary neural hypothesis concerned whether neural representations of outcome value would vary as a function of the partner contexts created by our social network manipulation. We constructed a random-effects general linear model (GLM) to examine the outcome phase of the

Table 1. Model parameters

Model	α (SE)	α gain (SE)	α loss (SE)	ß (SE)	Θ (SE)	Φ (SE)	AIC (SE)	ρR^2 (SE)	Recov. (SE)
Expected value	NA	NA	NA	0.98 (02)	NA	NA	106.34 (1.18)***	0.23 (0.02)	0.04 (0.03)
Expectation	0.02 (0.01)	NA	NA	0.35 (0.04)	NA	1.85 (0.15)	80.07 (4.18)*	0.42 (0.03)	0.89 (0.01)
Social value	0.21 (0.04)	NA	NA	0.59 (0.08)	2.64 (0.29)	NA	77.73 (3.80)	0.44 (0.03)	0.70 (0.01)
LG-RL	NA	0.79 (0.04)	0.10 (0.01)	0.37 (0.05)	NA	NA	80.50 (3.61)**	0.42 (0.03)	0.75 (0.01)
LG-RL partners	NA	F: 0.89 (0.04) C: 0.79 (0.05)†† CP: 0.61 (0.07)*	F: 0.08 (0.02) C: 0.09 (0.02) CP: 0.16 (0.03)†	0.31 (0.16)	NA	NA	84.71 (3.75)**	0.39 (0.03)	NA
Partner recip	0.23 (0.04)	NA	NA	0.63 (0.08)	F: 2.36 (0.33) C: 1.94 (0.27)* CP: 1.25 (0.19)**	NA	78.13 (4.08)	0.44 (0.03)	0.48 (0.01)

LG-RL Partners: *F*, Friend; C, Confederate; CP, Computer. Comparison Condition for LG-RL Partners: Friend.

$^{†}p < 0.10$, $^{††}p = 0.05$, $^{*}p < 0.05$, $^{**}p < 0.001$, $^{***}p < 0.0001$; Comparison Model: Social Value.

task by modeling reciprocation and defection experienced with each partner (six regressors). The variance associated with the decision phase was modeled with one regressor agnostic to partner and decision type. Six motion parameters, missed trials, and outcomes of keep decisions were included as confound regressors. Regressors of interest, missed trials, and keep decision/outcome regressors were *z*-transformed at the single-participant level and convolved with a double-gamma hemodynamic response function. A whole-brain probabilistic group mask excluding the skull based on the functional coverage of the data collected in our sample was applied to all whole-brain analyses.

Based on previous work (Fareri et al., 2012b) in which we observed social network modulation of BOLD responses during receipt of monetary rewards, our a priori neural hypothesis was that enhanced BOLD responses would emerge to reciprocation from a close friend compared with other outcomes. We tested this via a balanced contrast of friend reciprocate > all other outcomes. For completeness, a contrast of friend defect > all other outcomes was also conducted.

To explore whether participants were learning and adapting expected value via prediction error at the neural level, we constructed an additional random-effects GLM using parameters derived from our computational model at the whole-brain level. This GLM included four regressors of interest, including one general decision phase regressor, one general outcome phase regressor, and two regressors coding for model-derived parameters reflecting the probability of sharing on each trial and trial-by-trial prediction error values. Probability values were log transformed and then standardized within conditions across runs at the single-subject level. Prediction error values were standardized across runs and conditions at the single-subject level. All regressors of interest were convolved with a double-gamma hemodynamic response function. Six motion parameters were also included as regressors of no interest in this model. A whole-brain probabilistic mask excluding the skull based on functional coverage of the data collected in our sample was applied for group whole-brain analysis. We were primarily interested in determining what neural regions correlated with prediction error learning. At the group level, we conducted a simple *t* test of the parametric prediction error regressor (vs implicit baseline) to highlight regions linearly tracking prediction error across participants.

All statistical parametric maps (SPMs) were set to an initial uncorrected height threshold of $p < 0.001$. We corrected for multiple comparisons at the whole-brain level using the cluster level statistical estimator in BrainVoyager. This method of correction (Forman et al., 1995; Goebel et al., 2006) runs a series of Monte Carlo simulations on a given SPM to determine the likelihood that observed clusters of activation are significant and not false positives, resulting in a whole-brain corrected threshold of $p < 0.05$. All SPMs were corrected to a threshold of 17 contiguous voxels (459 mm^3).

Brain-model relationships. We tested whether the observed neural activation to reciprocation was related to the computationally derived social reward bonus parameters from the social value model. Mean parameter estimates from the whole-brain contrast of friend reciprocation > all other outcomes were extracted from resulting clusters in the SPM. We examined whether parameter estimates reflecting reciprocation for each partner predicted model-derived bonus parameters with a mixed-effects regression using the lmerTest and LME packages in R.

Results

Social closeness is associated with increased ratings of trustworthiness

Subjective ratings of social closeness (IOS; Aron et al., 1992) and partner trustworthiness were collected and analyzed as a manipulation check regarding social network. A one-way repeated-measures ANOVA on MRI participants' responses on the IOS revealed a significant main effect of partner ($F_{(1.62,\ 40.47)} = 25.06$, $p < 0.001$): participants reported increased levels of closeness with their friends compared with both the confederate ($t_{(25)} = 7.70$, $p < 0.001$), and the computer ($t_{(25)} = 6.37$, $p < 0.001$). A 3 (partner) × 2 (time) repeated-measures ANOVA on participants' presession and postsession trustworthiness ratings (Fig. 1*b*) revealed a significant main effect of partner ($F_{(1.63,\ 40.77)} = 37.91$, $p < 0.001$) such that MRI participants rated their friends as significantly more trustworthy than both the confederate ($t_{(25)} = 8.67$, $p < 0.00001$) and the computer ($t_{(25)} = 7.87$, $p < 0.00001$). Participant ratings of confederate trustworthiness were approaching a trend toward being higher than those of the computer ($t_{(25)} = 1.67$, $p = 0.11$). We did not observe a main effect of time ($F_{(1,25)} = 0.61$, $p > 0.40$) or an interaction ($F_{(2,50)} = 1.62$, $p > 0.20$).

Social network modulates collaborative decisions

The key behavioral measure of interest was whether participants' decisions to trust would vary as a function of the partner context created by our social network manipulation. A mixed-effects logistic regression of partner context on decision (i.e., share/keep) revealed that participants were more likely to share with their close friend ($\beta = 1.13$, SE $= 0.22$, $z = 5.24$, $p < 0.0001$) and the confederate ($\beta = 0.71$, SE $= 0.21$, $z = 3.35$, $p < 0.001$) compared with the computer (Fig. 1*c*). Participants were also more likely to share with their close friend compared with the confederate ($\beta = 0.42$, SE $= 0.14$, $z = 2.90$, $p < 0.005$). Critically, this behavioral pattern was observed despite all partners reciprocating at the same 50% rate during the trust game. Corroborating participants' decision patterns were the reaction time data, which indicated more rapid responses when choosing to share with a close friend ($\beta = -0.15$, SE $= 0.042$, $t = -3.47$, $p < 0.001$) and the confederate ($\beta = -0.11$, SE $= 0.04$, $t = -2.72$ $p < 0.01$) compared with the computer.

Computation of social value drives collaboration

Our primary hypothesis concerned whether participants' decisions to collaborate with a close friend are driven by a strong prior

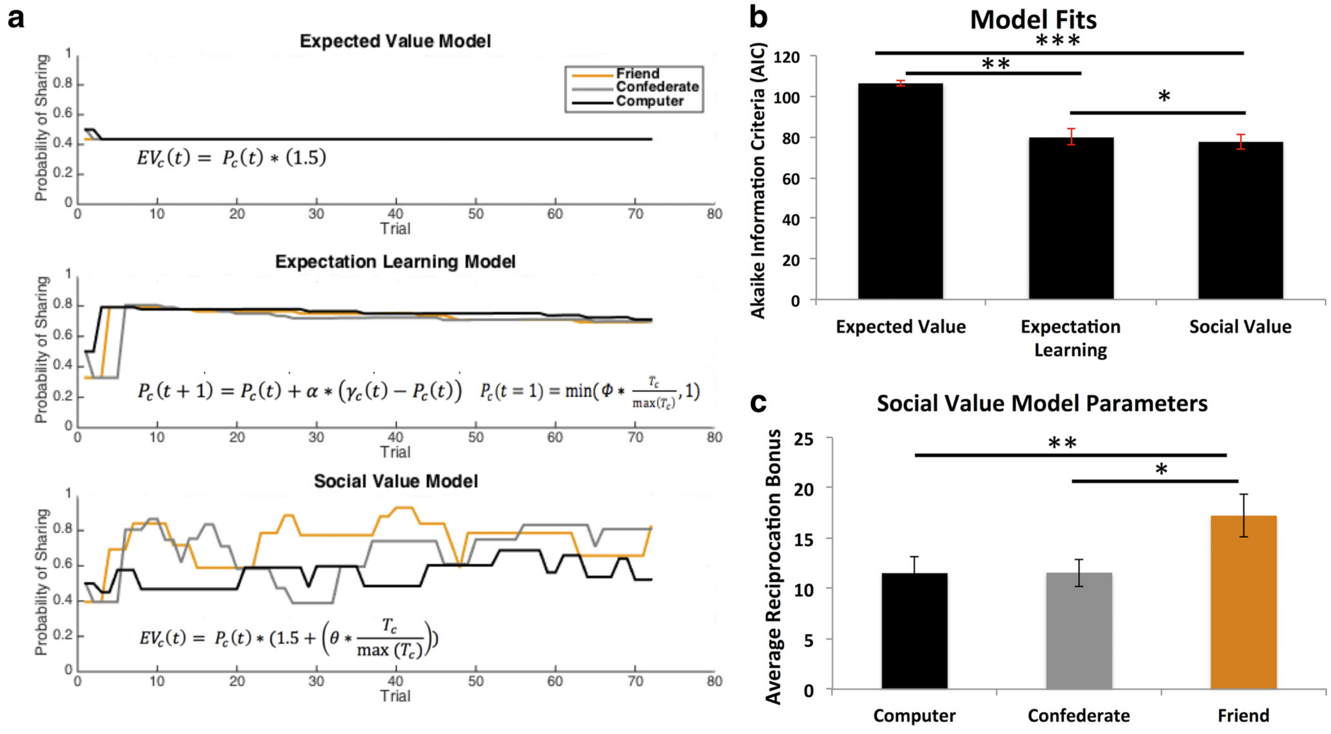

Figure 2. Computational model results. ***a***, Model simulations of the likelihood of sharing with a partner on a given trial for one randomly selected participant's experimental data. We compared the ability of the social value model (bottom) to explain collaborative behavior with a standard expected value model (top), which assumes that participants maximize expected value based solely on self-interested financial value and a 50% reinforcement rate, and an expectation-learning model incorporating strong priors, in which participants update their beliefs about partner reciprocation from both positive and negative outcomes (middle). ***b***, Average model fits penalizing for the number of free parameters using the Akaike Information Criteria (AIC). The social value model provided the best fit to participant behavior ($n = 26$). ***c***, Average social value bonus (theta * normalized trustworthiness ratings) for each partner. ***$p < 0.0001$; **$p < 0.001$; *$p < 0.05$ (±SEM).

expectation of reciprocation or by an enhanced social value to reciprocation. Nonparametric Wilcoxon signed-rank tests (Fig. 2*b*, Table 1) revealed that the social value model fit participants' data significantly better than the expected value model ($z = 4.46$, $p < 0.00001$) and the expectation-learning model ($z = 1.97$, $p < 0.05$). Parameter recovery exercises demonstrated that the number of trials did not adversely affect parameter estimation because we were able to recover the parameters used to simulate data in the social value and expectation-learning models with a relatively high accuracy (Table 1).

Given that the social value model provided the best fit to participants' data, we further investigated whether the social reward bonus parameters computed in this model varied as a function of partner. A repeated-measures ANOVA (Fig. 2*c*) revealed a significant effect of partner ($F_{(2,50)} = 12.90$, $p < 0.0001$) on social reward bonus parameters computed in the social value model such that participants assigned increased social value to reciprocation from a close friend (in-network) compared with an out-of-network ($t_{(25)} = 4.84$, $p < 0.0001$) or nonsocial ($t_{(25)} = 4.62$, $p < 0.001$) partner. No significant differences emerged between the social reward bonus attributed to the out-of-network partner and the computer ($t_{(25)} = 0.03$, $p > 0.97$).

Control models support social value model

We ran three additional control models to evaluate whether participants' decisions were driven by a general other-regarding preference (partner reciprocation value model), a sensitivity to reciprocation compared with defection generally (loss–gain RL model) or preferential discounting of negative outcomes (i.e., defection) experienced with their friends compared with other partners (loss–gain RL partners model).

First, the social value model and partner reciprocation value model did not significantly differ in terms of model fit ($z = 1.13$, $p = 0.26$). Importantly, estimating a separate theta for each partner in the partner reciprocation value model revealed a significant effect of partner ($F_{(2,50)} = 12.32$, $p < 0.0001$) such that sharing with a friend carries increased value compared with the confederate ($t_{(25)} = 1.79$, $p < 0.05$, one-tailed) or computer ($t_{(25)} = 4.68$, $p < 0.00005$, one-tailed). Although this result is not inconsistent with the preference for efficiency hypothesis, we believe that it is more consistent with the social value model because the degree of concern for efficiency was moderated by whether the partner was a friend or stranger.

Second, the social value model fit participants' data significantly better than the loss–gain RL model ($z = 3.14$, $p < 0.002$). In addition, the supplemental loss–gain RL partners model revealed that, although we observed a main effect of partner in how participants learned from negative outcomes in the loss–gain RL partners model ($F_{(2,50)} = 4.01$, $p < 0.05$), this was primarily driven by enhanced learning rates for the computer compared with the friend ($t_{(25)} = 2.43$, $p < 0.05$) and the confederate ($t_{(25)} = 1.94$, $p = 0.06$, trend). We found no evidence that participants were selectively discounting learning from negative outcomes with a friend compared with a confederate ($t_{(25)} = -0.53$, $p > 0.5$). Interestingly, a main effect of partner also emerged in the domain of positive outcomes ($F_{(2,50)} = 3.25$, $p < 0.0001$) because participants appeared to weight positive outcomes resulting from interactions with friends more strongly than the confederate ($t_{(25)} = 2.01$, $p = 0.06$, trend) or the computer ($t_{(25)} = 3.91$, $p < 0.001$), which is also consistent with the predictions of the social value model. Parameter recovery exercises demonstrate that the number of trials did not adversely affect parameter estimation because we were able to recover the parameters used to simulate data in the partner reciprocation value, and loss–gain RL models (Table 1).

Together, these results suggest that increased decisions to trust a close friend may be driven by the enhanced social value of in-network reciprocation as opposed to a general preference for efficiency in social interactions or selectively ignoring defection from a friend (e.g., compared with the confederate).

Social value recruits neural circuitry of reward

After our modeling results, we next sought to determine whether a social value signal would emerge in the brain corresponding to reciprocation from a close friend. A balanced contrast of friend reciprocate > all other outcomes revealed robust activation within putative reward circuitry (Fig. 3*a*–*c*, Table 2) including ventral striatum bilaterally (left: $x, y, z = -19, 7, -6$; right: $x, y, z = 11, 16, 3$), and medial prefrontal cortex (BA9/10: $x, y, z = -1, 55, 15$). No regions demonstrated stronger activation to friend defect > all other outcomes in a separate analysis (Table 3). In addition, we correlated model-derived trial-by-trial prediction error values at the whole-brain level and replicated our previous finding that the ventral striatum ($x, y, z = -10, 1, -9$) tracks a prediction error learning signal within a social context (Fareri et al., 2012a; Fig. 4, Table 4). Together, these results illustrate that the ventral striatum processes independently computations associated with both social value and socially relevant error signals used for updating beliefs about the likelihood of a partner reciprocating.

Magnitude of neural response to reward predicts model-derived social value

Although both the modeling and neuroimaging results indicate increased social value for reciprocation from close friends compared with other partners, these results do not imply a direct link between the model-derived value and the brain regions associated with social value. Therefore, we next examined the relationship between the model-derived bonus values for each condition and the outcome-related BOLD activation. Mean parameter estimates were extracted from the bilateral ventral striatal clusters defined by the contrast of friend reciprocate > all other conditions. A mixed-effects regression demonstrated that the average ventral striatal response at outcome for reciprocation significantly predicted the model-derived outcome bonus parameters ($\beta = 13.74$, SE = 6.79, $t = 2.02$, $p < 0.05$; Fig. 3*e*). The same analysis using extracted mean parameter estimates from the medial prefrontal cortex cluster also revealed a significant predictive relationship ($\beta = 10.71$, SE = 5.27, $t = 2.03$, $p < 0.05$; Fig. 3*d*) and no relationship for a visual cortex cluster ($\beta = -0.17$, SE = 0.57, $t = -0.29$, $p > 0.70$), suggesting some degree of specificity for this computation within putative reward circuitry.

Association between social closeness and neural response to reward

Based on our previous work demonstrating an association between social closeness and reward-related ventral striatal activation during a shared monetary reward task (Fareri et al., 2012b), we probed relationships between social closeness with a partner and neural responses to reciprocation in the present task. Mixed-effects linear regression revealed that self-reported social closeness with a partner significantly predicted BOLD responses to reciprocation in both ventral striatum ($\beta = 0.02$, SE = 0.006, $t = 3.39$, $p < 0.005$) and mPFC ($\beta = 0.03$, SE = 0.007, $t = 3.95$, $p < 0.0005$). These results further suggest that interpersonal aspects of a relationship can affect neural representations of reward value during social interactions.

Discussion

Collaboration facilitates the establishment and maintenance of trusting, meaningful personal relationships. This study sought to characterize the computational and neural mechanisms facilitating collaborative decisions. We hypothesized that people: (1) act to maximize their expected value, (2) update their expectations about partner trustworthiness using reinforcement learning, and (3) receive social value from positive interactions that drives future collaborations. To test these formal hypotheses, we manipulated the social network status of participants' partners (e.g., close friend or stranger) in the context of a repeated trust game. An important design feature of our study is that all partners reciprocated with a fixed 50% probability, ensuring that behavioral differences would primarily be due to factors such as the subjective quality of a relationship with one's interaction partners or prior expectations (e.g., perceived partner trustworthiness). Consistent with this idea, participants trusted their close friend more than a stranger or a computer despite equal reinforcement rates, a phenomenon that was driven by the computation of higher social reward for reciprocation from a close friend. Furthermore, we validated our social value model of collaboration by showing a selective relationship between model-derived outcome bonus parameters and both ventral striatal and mPFC activation at outcome and ruled out alternative explanations such as a strong prior expectation of reciprocation. The combination of these behavioral, computational, and imaging results highlights social value as the mechanism driving interpersonal collaboration.

Decisions to engage in collaborative behavior can be influenced strongly by expectations about others. Expectations for those with whom we have no experience may be based on assumed knowledge about their social group (Stanley et al., 2011, 2012), perceived trustworthiness based on facial characteristics (van 't Wout and Sanfey, 2008), or instructed knowledge regarding their moral character (Delgado et al., 2005). Such expectations about new partners can interact with actual experience and be updated dynamically (Chang et al., 2010; Fareri et al., 2012a). Our data indicate that expectations appear to guide behavior less so when considering established relationships. Instead, the context created by close relationships suggests that collaboration is driven by a social value signal in which reciprocation is more valued when coming from a close friend compared with someone unknown. This enhanced social value may serve to reaffirm and maintain the relationship, thus satisfying a need to belong (Baumeister and Leary, 1995).

Our neural results synthesize two hypotheses that have been posited regarding the role of the ventral striatum in collaborative interactions. The first hypothesis is that ventral striatal activity supports the process of learning a partner's reputation during iterated trust games (Delgado et al., 2005; King-Casas et al., 2005; Fareri et al., 2012a). Ventral striatal activation at the time of social outcome has been shown to correlate directly with model-derived prediction error learning signals (Fareri et al., 2012a) and appears to propagate to the earliest predictor of trust after repeated reciprocation, which is consistent with temporal difference learning (King-Casas et al., 2005; for review see Kishida and Montague, 2012). Consistent with this hypothesis, our model proposed that, within collaborative interactions, expectations about a partner will be updated based on experienced outcomes

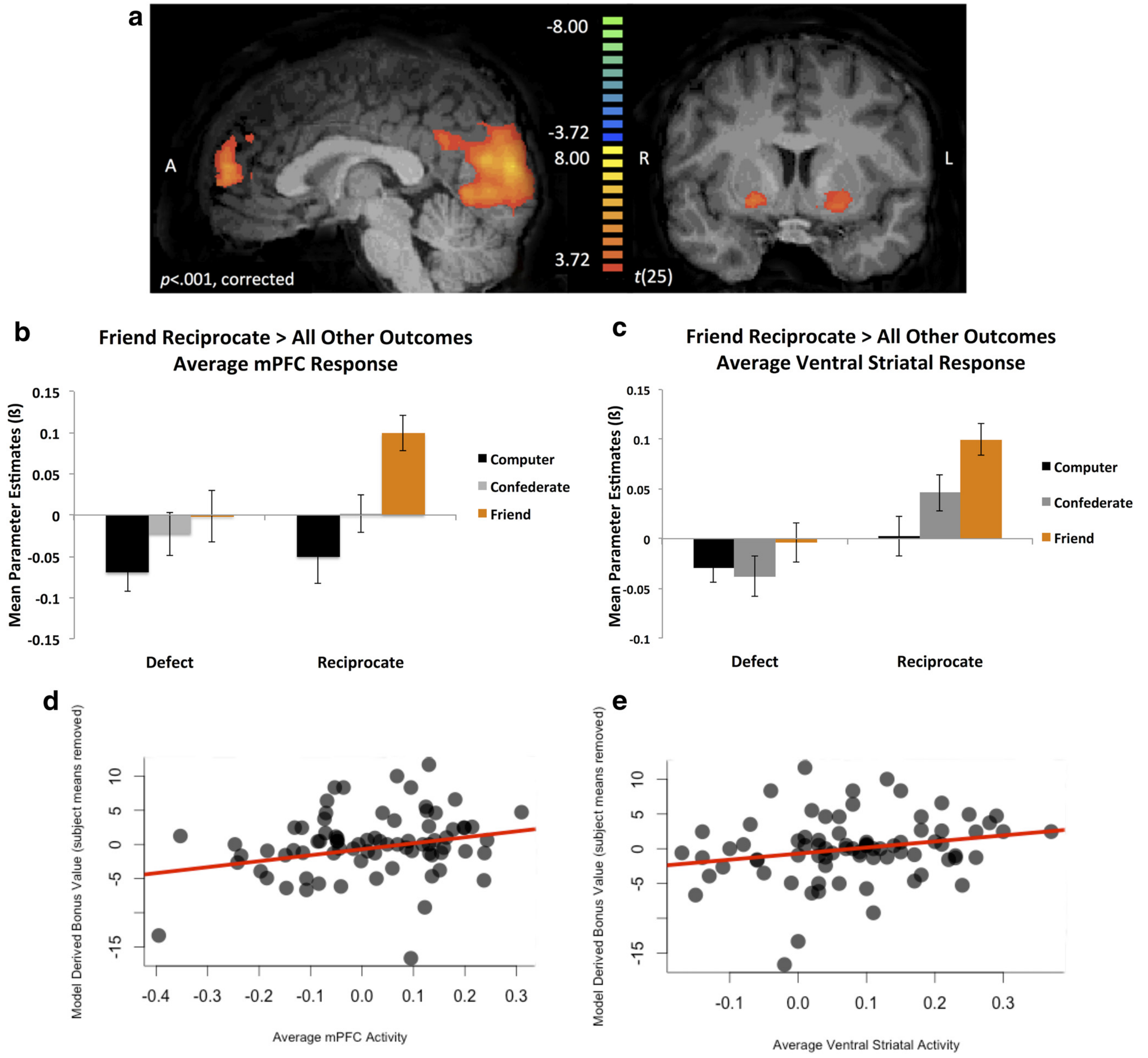

Figure 3. Neural representations of social value. ***a***, A whole-brain balanced contrast of friend reciprocate > all other outcomes revealed significant clusters of activation within putative reward circuitry, including ventral striatum, bilaterally (left: $x, y, z = -19, 7, -6$; right: $x, y, z = 11, 16, 3$) and medial prefrontal cortex (BA9/10: $x, y, z = -1, 55, 15$). Results are depicted at $p < 0.001$, whole-brain cluster corrected to $p < 0.05$ ($n = 26$). ***b***, ***c***, Plot of extracted mean parameter estimates depicting the average mPFC ($x, y, z = -1, 55, 15$; ***b***) and ventral striatal (left: $x, y, z = -19, 7, -6$; right: $x, y, z = 11, 16, 3$; ***c***) response identified via the contrast of friend reciprocate > all other outcomes (±SEM). ***d***, ***e***, Scatterplots depicting the relationship between model-derived bonus values (theta * normalized trustworthiness ratings) and average mPFC (***d***) and ventral striatal (***e***) activity. Participants' means of the bonus parameter have been removed to be consistent with the mixed-effects regression analysis.

Table 2. Friend reciprocate > all other conditions (balanced)

Region of activation (peak)	Brodmann area	Laterality	Talairach coordinates (peak) *x*	*y*	*z*	*t*-stat	# Voxels (mm^3)
Fusiform gyrus	BA19	R	23	−56	−9	8.39	51654
Caudate nucleus		R	11	16	3	5.21	460
Medial frontal gyrus	BA9/10	L	−1	55	15	5.74	4345
Ventral striatum		L	−19	7	−6	6.03	871

Table 3. Friend defect > all other outcomes (balanced)

Region of activation (peak)	Brodmann area	Laterality	Talairach coordinates (peak) *x*	*y*	*z*	*t*-stat	# Voxels (mm^3)
Postcentral gyrus	BA43	R	65	−17	18	−6.12	1718
Superior frontal gyrus	BA10	R	38	49	18	−4.97	670
Cuneus	BA18	R	5	−92	9	−5.96	7209

via RL (Rescorla and Wagner, 1972). We indeed found that participants dynamically updated expectations via prediction error learning because behavior was better explained by models accounting for learning than those positing no learning and model-derived prediction errors significantly correlated with ventral striatal activation. This replicates previous findings implicating the ventral striatum in signaling prediction errors (Garrison et al., 2013) and in coding social learning signals (Fareri et al., 2012a; Fouragnan et al., 2013).

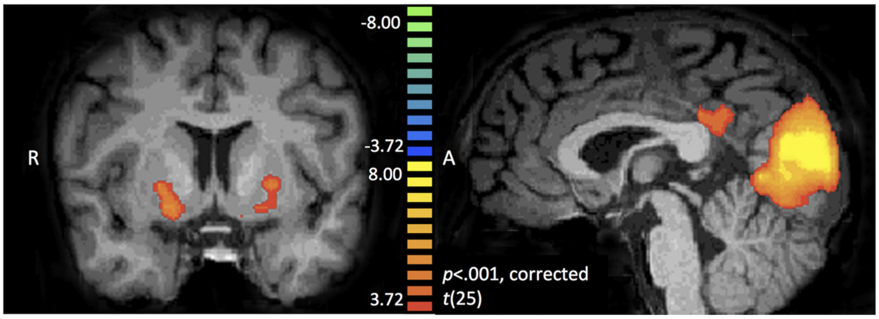

Figure 4. Neural correlates of prediction error. A whole-brain analysis using model-derived prediction error values from the social value model showed enhanced prediction-error-related activation in the bilateral striatum, including left ventral striatum ($x, y, z = -10, 1, -9$) and right caudate nucleus extending to the ventral striatum ($x, y, z = 14, 16, 3$) and posterior cingulate ($x, y, z = -1, -35, 30$), among other regions. Results are depicted at $p < 0.001$ whole-brain cluster corrected to $p < 0.05$.

Table 4. Regions tracking prediction error

			Talairach coordinates (peak)				
Region of activation (peak)	Brodmann area	Laterality	x	y	z	t-statistic	# Voxels (mm^3)
Precuneus	BA19	R	29	−62	39	5.29	2066
Cuneus	BA18/17	L	−4	−86	12	10.13	67230
Caudate nucleus		R	14	16	3	6.04	1477
Cingulate gyrus	BA31	L	−1	−35	30	4.98	782
Ventral striatum		L	−10	1	−9	5.35	546
Middle occipital gyrus	BA37	L	−46	−68	−6	6.03	1011

A second hypothesis is that ventral striatal responses to reciprocated trust reflect a reward signal that drives future collaborative decisions. This hypothesis has primarily been observed in Prisoner's dilemma (Rilling et al., 2002, 2004) and trust (Phan et al., 2010) games. In these studies, the ventral striatum codes for outcomes achieved via mutual collaboration (Rilling et al., 2002), specifically with human partners (Rilling et al., 2002, 2004) who have a reputation for reciprocity (Phan et al., 2010). Our model proposes that participants receive value from reciprocation, but that this value is composed of both self-interested financial value and social value. Consistent with this second hypothesis, we observed increased activation in the ventral striatum to reciprocation from a close friend.

The present study furthers our understanding of the mechanisms supporting collaboration by building on these hypotheses in two critical ways. First, our findings synthesize proposed roles of the ventral striatum in reputation learning and coding reciprocated trust as reward, suggesting that both processes are necessary to support collaboration simultaneously. Second, we highlight a social value model that accounted for participant behavior. The magnitude of the social reward bonus signal predicted by the model was associated with the strength of ventral striatal and mPFC BOLD responses to reciprocation. In conjunction with results indicating that the degree of social closeness between partners was also associated with the neural response to reciprocation, our findings provide a mechanistic account of reward circuit function during collaborative interactions.

The mPFC is known to play a significant role in processing socially relevant information, and signals reflecting valuation processes (for review, see Amodio and Frith, 2006; Bartra et al., 2013; Fareri and Delgado, 2014b). Representations of the self and of close others recruit mPFC (Heatherton et al., 2006; Mitchell et al., 2006; Krienen et al., 2010) and, within social interactions, distinct regions of mPFC represent cooperative human partners (McCabe et al., 2001), volatility of the social environment (Behrens et al., 2008), and strategic processing of outcomes/mentalizing about another's actions (Hampton et al., 2008; Fareri and Delgado, 2014a). Our results showing mPFC BOLD responses to reciprocation predicting model-derived outcome bonus parameters merges with the extant literature by demonstrating that, within collaborative interactions, mPFC may work with the ventral striatum to perform computations that incorporate both social and outcome-related information that drive future collaborations.

An alternative interpretation of our results is that participants ignore negative outcomes (e.g., partner defection) when they have a strong prior feeling about a partner, failing to update their expectations in the face of inconsistent information. We believe that this is an unlikely interpretation of our data. Because this is essentially a variation on prediction error learning, it would follow that the ventral striatal activity should be strongest for the stranger condition after reciprocation, lowest for the friend condition when they choose to keep, and no response when their friend chose to share because this would have be predicted previously by their prior expectation. There is no evidence of this pattern of activity in our study. Further, although our social value model cannot test directly whether participants weighted negative outcomes differentially as a function of partner, the behavioral analysis, which separately modeled learning rates for positive and negative outcomes as a function of partner, did not reveal a discounting of negative outcomes unique to the close friend condition. It is possible that learning about reputations via descriptive means (Delgado et al., 2005; Fouragnan et al., 2013) modulates striatal signaling during collaborative interactions via prefrontal mechanisms as in nonsocial contexts, decreasing the reliance on trial-by-trial updating (Li et al., 2011). Although we cannot completely rule out this possibility, we believe that our results speak more to the assignment of added social value to reciprocation from a close friend, which subsequently drives collaborative behavior in close relationships, as opposed to a unique discounting of negative outcomes experienced with a close friend.

A second alternative interpretation of our data is that participants' decisions to collaborate are driven by general concerns for efficiency or social welfare in interactions regardless of the influence of a relationship. Individuals are often motivated by other-regarding preferences, acting to minimize situations of unfairness (Charness and Rabin, 2002; Hsu et al., 2008; Tricomi et al., 2010) and ensuring more equitable distributions of outcomes. Further, the striatum has been implicated in coding a utility signal representing both efficiency and social welfare during social decision making (Hsu et al., 2008). Our partner reciprocation value model provides some support against solely an efficiency hypothesis because theta was significantly different for each partner and greatest for the friend. Our data are thus consistent with a framework of other-regarding preferences motivating behavior in social interactions, but future studies may aim to disentangle contributions of concerns for efficiency, other-regarding preferences, and the social value of a relationship as motivation for collaborative decisions.

The combination of a neuroeconomic approach to investigating the effect of social network on decision making with computational modeling techniques is an exciting direction for social neuroscience and one that has garnered much recent interest (Stanley and Adolphs, 2013). To date, however, only a handful of investigations have used this technique in the social domain (Behrens et al., 2008; Hampton et al., 2008; Chang and Sanfey, 2013; Fareri et al., 2012a; Fouragnan et al., 2013; Xiang et al., 2013). This approach allows us to test mechanistic hypotheses regarding social behavior and motivates specific predictions for how such socially relevant computations may be represented at the neural level. Importantly, this approach allowed us to synthesize interpretations in the literature regarding the role of the ventral striatum in collaborative behavior and provide strong support for the underlying psychological and neural mechanisms sustaining decisions to cooperate within close interpersonal relationships.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716 –723. CrossRef

Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7:268 –277. CrossRef Medline

Aron A, Aron EN, Smollan D (1992) Inclusion of other in the self scale and the structure of interpersonal closeness. J Person Soc Psych 63: 596 – 612.

Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76:412– 427. CrossRef Medline

Bates D, Maechler M, Walker S, Haubo Bojesen Christense R, Singmann H (2014) Linear mixed-effects modles using Eigen and S4. http://cran.r-project.org/web/packages/lme4/lme4.pdf.

Baumeister RF, Leary MR (1995) The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychological Bulletin 117:497–529. CrossRef Medline

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245–249. CrossRef Medline

Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. Games and Economic Behavior 10:122–142. CrossRef

Camerer C, Ho TH (1999) Experience-weighted attraction learning in normal form games. Econometrica 67:827– 874.

Chang LJ, Sanfey AG (2013) Great expectations: neural computations underlying the use of social norms in decision-making. Soc Cogn Affect Neurosci 8:277–284. CrossRef Medline

Chang LJ, Doll BB, van't Wout M, Frank MJ, Sanfey AG (2010) Seeing is believing: trustworthiness as a dynamic belief. Cogn Psychol 61:87–105. CrossRef Medline

Charness G, Rabin M (2002) Understanding social preferences with simple tests. The Quarterly Journal of Economics 117:1–53. CrossRef

Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systmes of reward during the trust game. Nat Neurosci 8:1611–1618. CrossRef Medline

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ (2009) Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. Brain Res 1299:74 –94. CrossRef Medline

Fareri DS, Delgado MR (2014a) Differential reward responses during competition against in- and out-of-network others. Soc Cogn Affect Neurosci 9:412– 420. CrossRef Medline

Fareri DS, Delgado MR (2014b) The importance of social rewards and social networks in the human brain. Neuroscientist 20:387– 402. CrossRef Medline

Fareri DS, Chang LJ, Delgado MR (2012a) Effects of direct social experience on trust decisions and neural reward circuitry. Front Neurosci 6:148. Medline

Fareri DS, Niznikiewicz MA, Lee VK, Delgado MR (2012b) Social network modulation of reward-related signals. J Neurosci 32:9045–9052. CrossRef Medline

Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics 114:817– 868. CrossRef

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magnetic resonance medicine. Magn Reson Med 33:636 – 647. CrossRef Medline

Fouragnan E, Chierchia G, Greiner S, Neveu R, Avesani P, Coricelli G (2013) Reputational priors magnify striatal responses to violations of trust. J Neurosci 33:3602–3611. CrossRef Medline

Garrison J, Erdeniz B, Done J (2013) Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310. CrossRef Medline

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp 27:392– 401. CrossRef Medline

Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A 105:6741– 6746. CrossRef Medline

Heatherton TF, Wyland CL, Macrae CN, Demos KE, Denny BT, Kelley WM (2006) Medial prefrontal activity differentiates self from close others. Soc Cogn Affect Neurosci 1:18 –25. CrossRef Medline

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65–70.

Hsu M, Anen C, Quartz SR (2008) The right and the good: distributive justice and neural encoding of equity and efficiency. Science 320:1092–1095. CrossRef Medline

Jones RM, Somerville LH, Li J, Ruberry EJ, Libby V, Glover G, Voss HU, Ballon DJ, Casey BJ (2011) Behavioral and neural properties of social reinforcement learning. J Neurosci 31:13039 –13045. CrossRef Medline

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: reputation and trust in a two-person economic exchange. Science 308:78 – 83. CrossRef Medline

Kishida KT, Montague PR (2012) Imaging models of valuation during social interaction in humans. Biol Psychiatry 72:93–100. CrossRef Medline

Krienen FM, Tu PC, Buckner RL (2010) Clan mentality: evidence that the medial prefrontal cortex responds to close others. J Neurosci 30:13906 –13915. CrossRef Medline

Kuznetsova A, Brockhoff PB, Haubo Bojesen Christense R (2014) Tests for random and fixed effects for linear mixed effect models (lmer objects of lme 4 package). http://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf.

Li J, Delgado MR, Phelps EA (2011) How instructed knowledge modulates the neural systems of reward learning. Proc Natl Acad Sci U S A 108:55– 60. CrossRef Medline

McCabe K, Houser D, Ryan L, Smith V, Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci U S A 98:11832–11835. CrossRef Medline

Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50: 655– 663. CrossRef Medline

Phan KL, Sripada CS, Angstadt M, McCabe K (2010) Reputation for reciprocity engages the brain reward center. Proc Natl Acad Sci U S A 107: 13099 –13104. CrossRef Medline

Rabin M (1993) Incorporating fairness into game theory and economics. The American Economic Review 83:1281–1302.

Rescorla R, Wagner A (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcment and non reinforcement (Black A, Prokasy W, eds). New York: Appleton-Century-Crofts.

Rice W (1989) Analyzing tables of statistical tests. Evolution 43:223–225. CrossRef

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. Neuroreport 15:2539 –2543. CrossRef Medline

Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002) A neural basis for social cooperation. Neuron 35:395– 405. CrossRef Medline

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599. CrossRef Medline

Simpson JA (2007) Psychological foundations of trust. Current Directions in Psychological Science 16:264 –268. CrossRef

Stanley DA, Adolphs R (2013) Toward a neural basis for social behavior. Neuron 80:816 – 826. CrossRef Medline

Stanley DA, Sokol-Hessner P, Banaji MR, Phelps EA (2011) Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proc Natl Acad Sci U S A 108:7710 –7715. CrossRef Medline

Stanley DA, Sokol-Hessner P, Fareri DS, Perino MT, Delgado MR, Banaji MR, Phelps EA (2012) Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. Philos Trans R Soc Lond, B, Biol Sci 367:744 –753. CrossRef Medline

Sutton R, Barto A (1998) Reinforcement learning. Cambridge, MA: MIT.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: George Thieme Verlag.

Tricomi E, Rangel A, Camerer CF, O'Doherty JP (2010) Neural evidence for inequality-averse social preferences. Nature 463:1089 –1091. CrossRef Medline

Uchino BN (2009) Understanding the links between social support and physical health. Perspectives in Psychological Science 4:236 –255. CrossRef

van't Wout M, Sanfey AG (2008) Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. Cognition 108:796 – 803. CrossRef Medline

Xiang T, Lohrenz T, Montague PR (2013) Computational substrates of norms and their violations during social exchange. J Neurosci 33:1099 – 1108a. CrossRef Medline