Associative learning of social value

Timothy E. J. Behrens[1,2]*, Laurence T. Hunt[1,2]*, Mark W. Woolrich[1] & Matthew F. S. Rushworth[1,2]

Our decisions are guided by information learnt from our environment. This information may come via personal experiences of reward, but also from the behaviour of social partners[1,2]. Social learning is widely held to be distinct from other forms of learning in its mechanism and neural implementation; it is often assumed to compete with simpler mechanisms, such as reward-based associative learning, to drive behaviour[3]. Recently, neural signals have been observed during social exchange reminiscent of signals seen in studies of associative learning[4]. Here we demonstrate that social information may be acquired using the same associative processes assumed to underlie reward-based learning. We find that key computational variables for learning in the social and reward domains are processed in a similar fashion, but in parallel neural processing streams. Two neighbouring divisions of the anterior cingulate cortex were central to learning about social and reward-based information, and for determining the extent to which each source of information guides behaviour. When making a decision, however, the information learnt using these parallel streams was combined within ventromedial prefrontal cortex. These findings suggest that human social valuation can be realized by means of the same associative processes previously established for learning other, simpler, features of the environment.

To compare learning strategies for social and reward-based information, we constructed a task in which each outcome revealed information both about likely future outcomes (reward-based information) and about the trust that should be assigned to future advice from a confederate (social information).

Twenty-four subjects performed a decision-making task requiring the combination of information from three sources (Fig. 1, Methods and Supplementary Information): (1) the reward magnitude of each option (generated randomly at each trial); (2) the likely correct response (blue or green) based on their own experience of rewards on each option; and (3) the confederate's advice, and how trustworthy the confederate currently was. When a new outcome was witnessed, subjects could use this single outcome to learn in parallel about the likely correct action, and the trustworthiness of the confederate.

The investigation resembles previous experiments that have compared animate and inanimate conditions in different trials or experiments[5,6]. Here, however, both sources of information were present on each trial outcome but the relevance of each was manipulated continuously allowing determination of both the functional magnetic resonance imaging (fMRI) signal and the behavioural influence associated with each source of information.

Optimal behaviour in this task requires the subject to track the probability of the correct action and the probability of correct advice independently, and to combine these two probabilities into an overall probability of the correct response (Supplementary Information). Computational models of reinforcement learning (RL) have had considerable success in predicting how such probabilities are tracked in learning tasks outside the social domain[7]. The simplest RL models integrate information over trials by maintaining and updating the expected value of each option. When new information is observed this value is updated by the product of the prediction error and the learning rate[7]. In our task, there are two dissociable prediction errors: the reward prediction error (actual reward − expected value), for learning about the correct option, and the confederate prediction error (actual − expected fidelity), for learning about the trustworthiness of the confederate. The optimal learning rate depends on the volatility of the underlying information source[8–10]. In volatile conditions, subjects should give more weight to recent information, using a fast learning rate. In stable conditions, subjects should weigh recent and historical information almost equally, using a slow learning rate. By ensuring that the correct option and the confederate's advice became volatile at different times, we ensured that the learning rate for these two sources of information varied independently. We used a Bayesian RL[8] model (Supplementary Information) to generate the optimal estimates of prediction error, volatility and outcome probability separately for each source of information (Fig. 1b–d).

We first sought to establish whether human behaviour matched predictions from the RL model. We used logistic regression to determine the degree to which subject choices were influenced by the optimally tracked confederate and outcome probabilities, and by the difference in reward magnitudes between options. Parameter estimates for all three information sources were significantly greater than zero, and there was no significant difference in the degree to which subjects used reward and social information to determine their behaviour (Fig. 1e). Furthermore there was no significant effect either of subjects blindly following confederate advice without learning its value, or of subjects assuming that the confederate would behave in the same way as the previous trial (Fig. 1e). Hence subjects were able to integrate the fidelity of the confederate over many trials in an RL-like fashion.

We then investigated whether the fMRI signal reflected the model's estimates of prediction error and volatility, for both social and reward information, when subjects witnessed new outcomes. In the reward domain, neural responses have been identified that encode these key parameters[8,11–16]. Dopamine neurons in the ventral tegmental area (VTA) code reward prediction errors[12,13,17]. Similar signals are reported in the dopaminoceptive striatum[11,18] and even in the VTA itself, when specialized strategies are used in human fMRI studies[19]. fMRI correlates of the learning rate in the reward domain have been reported in anterior cingulate sulcus (ACCs)[8]. If humans can learn from social information in a similar fashion, it should be possible to detect signals that co-vary with the same computational parameters, but in the social domain.

We observed blood-oxygen-level-dependent (BOLD) correlates of the confederate prediction error in dorsomedial prefrontal cortex (DMPFC) in the vicinity of the paracingulate sulcus, right middle temporal gyrus (MTG), and in the right superior temporal sulcus at the temporoparietal junction (STS/TPJ) (Fig. 2a). Equivalent signals

[1]FMRIB Centre, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. [2]Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.
*These authors contributed equally to this work.

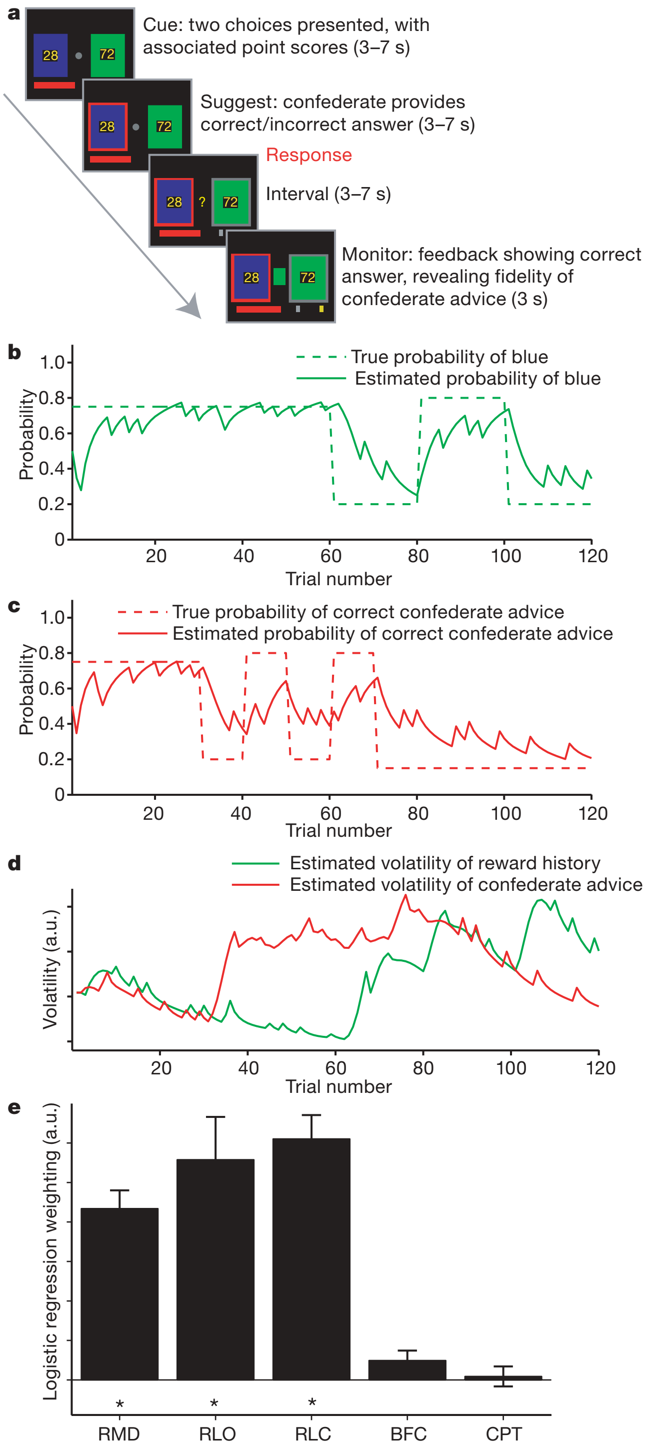

Figure 1 | Experimental task and behavioural findings. **a**, Experimental task (see Methods and Supplementary Information). Each trial consists of four phases. Subjects are presented with a decision (Cue), receive the advice (red square) of the confederate (Suggest) and respond using a button press (grey square). An 'Interval' period follows, before the correct outcome is revealed (Monitor). If the subject chooses correctly the red bar is incrementally increased by the number of points on the chosen option. **b**, **c**, Reward schedules for reward (**b**) and social (**c**) information. Dashed lines show the true probability of blue being correct (**b**) and the true probability of correct confederate advice (**c**). Each schedule underwent periods of stability and volatility. Solid lines show the model's estimate of the probabilities. **d**, Optimal model estimates of the volatility of reward (green) and social (red) information. **e**, Logistic regression on subject behaviour. Factors included were the reward magnitude difference between options (RMD); the outcome probability derived from the model using reward outcomes (RLO); the outcome probability derived from the model using confederate advice (RLC); the possibility that the subjects would blindly follow the confederate without learning (BFC); and the possibility that subjects would assume the confederate would behave as in the previous trial (CPT). The logistic regression analysis revealed significant effects only on RMD, RLO and RLC (asterisks). Error bars show s.e.m.; a.u., arbitrary units.

were present in the left hemisphere at the same threshold, but did not pass the cluster extent criterion; similar effects were also found bilaterally in the cerebellum (Supplementary Information). Notably, these regions showed a pattern of activation similar to known dopaminergic activity in reward learning[13], but for social information. Activity correlated with the probability of a confederate lie after the subject decision but before the outcome was revealed (a prediction signal). When the subjects observed the trial outcome, activity correlated negatively with this same probability, but positively with the event of a confederate lie (Fig. 2b). This signal reflects both components of a prediction error signal for social information: the outcome (lie or truth) minus the expectation (Fig. 2b). These signals cannot be influenced by reward prediction errors as the two types of prediction error were decorrelated in the task design. The presence of this prediction error signal in the brain is a prerequisite for any theory of an RL-like strategy for social valuation.

We performed a similar analysis for prediction errors on reward information (reward minus expected reward). We found a significant effect of reward prediction error in the ventral striatum (Fig. 2c), the ventromedial prefrontal cortex, and anterior cingulate sulcus (see Supplementary Information). As in the social domain, we observed significant effects of all three elements of the reward prediction error (Fig. 2d; see Supplementary Information for discussion).

As previously demonstrated[8], the volatility of action–outcome associations predicted BOLD signal in a circumscribed region of the ACCs (Fig. 3a). This effect varied across people such that those whose behaviour relied more on their own experiences (Supplementary Information) showed a greater volatility-related signal in this region (Fig. 3b). The volatility of confederate advice correlated with BOLD signal in a circumscribed region in the adjacent ACC gyrus (ACCg) (Fig. 3a). Subjects whose behaviour relied more on this advice showed greater signal change in this region (Fig. 3c). Notably, this double dissociation (reflected in a three-way interaction between area (ACCs versus ACCg), volatility type (social versus outcome) and degree of reliance on social ($F_{1,20} = 7.145$, $P = 0.015$) or experiential information ($F_{1,20} = 5.379$, $P = 0.031$) in an analysis of covariance) can be understood by reference to a dissociation in macaque monkeys. Selective lesions to ACCs but not ACCg impair reward-guided decision making in the reward domain[20]. In the social domain, male macaques will forego food to acquire information about other individuals[21,22]. Selective lesions to ACCg but not ACCs abolish this effect[23]. We found that BOLD signals in these two regions reflect the respective values of the same outcome for learning about the two different sources of information.

Learning about reward probability from vicarious and personal experiences recruits distinct neural systems, but subjects combine information across both sources when making decisions (Fig. 1e). A ventromedial portion of the prefrontal cortex (VMPFC) has been shown to code such an expected value signal for the chosen action[24,25] during decision making.

We computed two probabilities of reward on the subject's chosen option: one based only on experience and one based only on confederate advice. BOLD signal in the VMPFC was significantly correlated with both probabilities (Fig. 4a and Supplementary Fig. 4). However, there was subject variability in whether the VMPFC signal better reflected the reward probability based on outcome history or on social information. The extent to which the VMPFC data reflected each source of information (at the time of the decision) was predicted by the ACCs/ACCg response to outcome/social volatility (at the time when the outcomes were witnessed) (Fig. 4b, c).

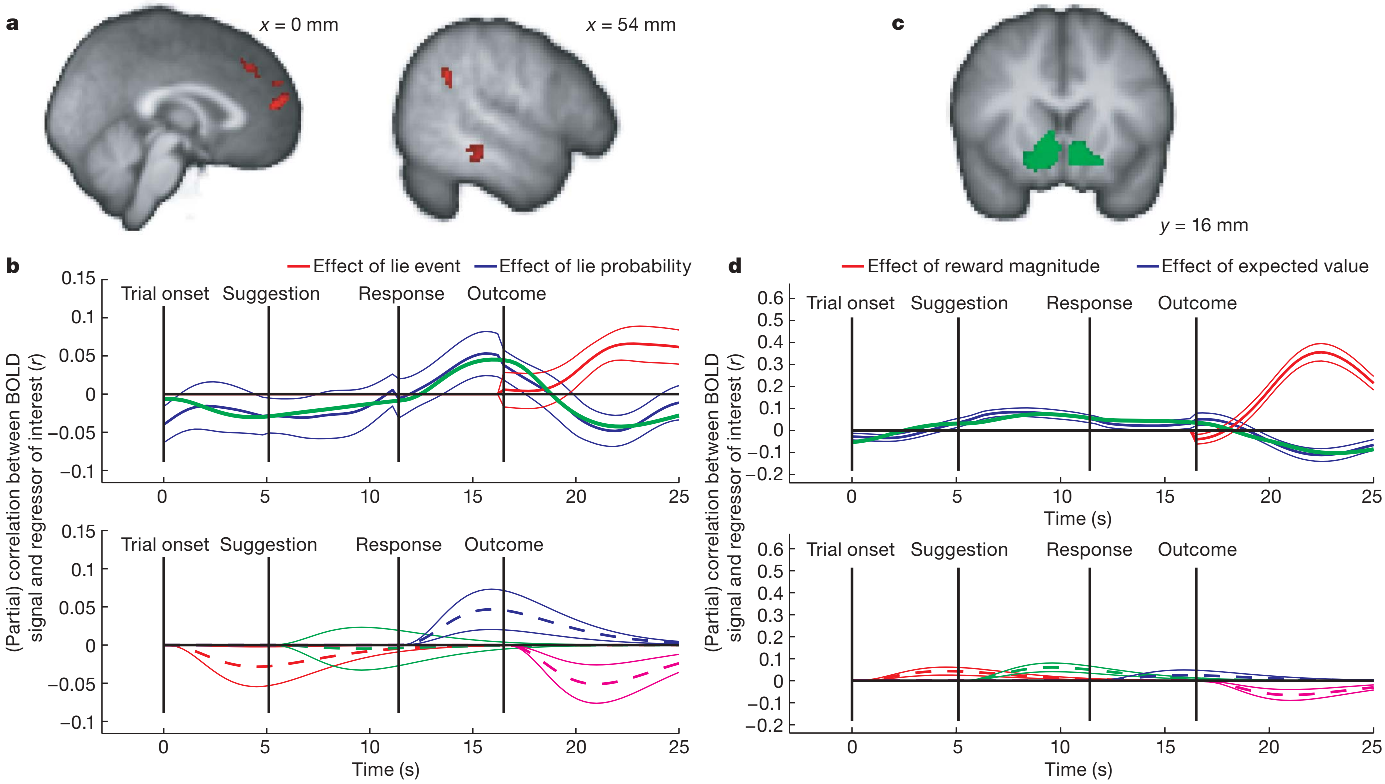

Figure 2 | Predictions and prediction errors in social and non-social domains. Time courses show (partial) correlations ± s.e.m. (See Supplementary Fig. 2.) **a**, Activation in the DMPFC, right TPJ/STS and MTG correlate with the social prediction error at the outcome (threshold set at $Z > 3.1$, cluster size >50 voxels). **b**, Deconstruction of signal change in the DMPFC. Similar results were found in the MTG and TPJ/STS. Top: following the outcome, areas that encode prediction error correlate positively with the outcome and negatively with the predicted probability. Red, effect size of the confederate lie outcome (1 for lie, 0 for truth); blue, effect size of the predicted confederate lie probability. To perform inference, we fit a haemodynamic model in each subject to the time course of this effect (that is, to the blue line). The green line in the top panel shows the mean overall fit of this haemodynamic model (for comparison with the blue line). Bottom: the effect of lie probability (blue line from top panel) is decomposed into a haemodynamic response function at each trial event (corresponding to the four colours in the bottom panel) (see Supplementary Fig. 2). Dashed and solid lines show mean responses ± s.e.m. Each region showed a significant positive effect of predicted confederate lie probability after the decision ($t_{22} = 1.96$ ($P < 0.05$), 1.73 ($P < 0.05$), 1.74 ($P < 0.05$) for DMPFC, MTG and TPJ/STS, respectively). Crucially, each brain region showed a significant negative effect of predicted confederate lie probability after the outcome ($t_{22} = 2.68$ ($P < 0.005$), 2.35 ($P < 0.05$), 3.27 ($P < 0.005$)). **c**, Ventral striatum is taken as an example of a number of regions revealed by the voxel-wise analysis of reward prediction error (threshold set at $Z > 3.1$, cluster size >100 voxels). **d**, Panels are exactly as in **b**, but coded in terms of reward and not in terms of confederate fidelity. The top panel shows the parameter estimate relating to the expected value of the trial (blue line) and, after the outcome, the parameter estimate relating to the magnitude of these rewards (red line). To test for prediction error coding, we again fit a haemodynamic model to the expectation parameter estimate (shown by the green line, for comparison with blue line). Bottom panel: the time course showed a significant positive effect during the time of the decision ($t_{22} = 3.32$ ($P < 0.002$)), and a significant negative effect after the trial outcome ($t_{22} = 2.50$ ($P < 0.05$)). (See Supplementary Information for further discussion.)

Here, we have shown that the weighting assigned to social information is subject to learning and continual update via associative mechanisms. We use techniques that predict behaviour when learning from personal experiences to show that similar mechanisms explain behaviour in a social context. Furthermore, we demonstrate fundamental similarities between the neural encoding of key parameters for reward-based and social learning. Despite using similar mechanisms, distinct anatomical structures code learning parameters in the two domains. However, information from both is combined in ventromedial prefrontal cortex when making a decision.

By comparing the two sources of information, we find that social prediction error signals similar to those reported in dopamine neurons for reward-based learning are coded in the MTG, STS/TPJ and DMPFC. BOLD signal fluctuations in these regions are often seen in social tasks[26,27], and in tasks which involve the attribution of motive to stimuli[28]. Such activations have been thought critical in studies of the theory of mind[28]. That these regions should code quantitative prediction and prediction error signals about a confederate lends more weight to the argument that social evaluation mechanisms are able to rely on simple associative processes.

A second crucial parameter in reinforcement learning models is the learning rate, reflecting the value of each new piece of information. In the context of reward-based learning, this parameter predicts BOLD signal fluctuations in the ACCs at the crucial time for learning[8]—a finding that is replicated here. We further demonstrate that the exact same computational parameter, in the context of social learning, predicts BOLD fluctuations in the neighbouring ACCg. This functional dissociation is mirrored by differences in the regions' anatomical connectivity. In the macaque monkey, connections with motor regions lie predominantly in ACCs[29], giving access to information about the monkey's own actions. Connections with visceral and social regions, including the STS, lie predominantly in ACCg[29], giving access to information about other agents. Nevertheless, that it is the same computational parameter that is represented in ACCs and ACCg suggests that parallel streams of learning occur within ACC for social and non-social information.

It has been suggested that VMPFC activity might represent a common currency in which the value of different types of items might be encoded[25,30]. Here we show that the same portion of the VMPFC represents the expected value of a decision based on the combination

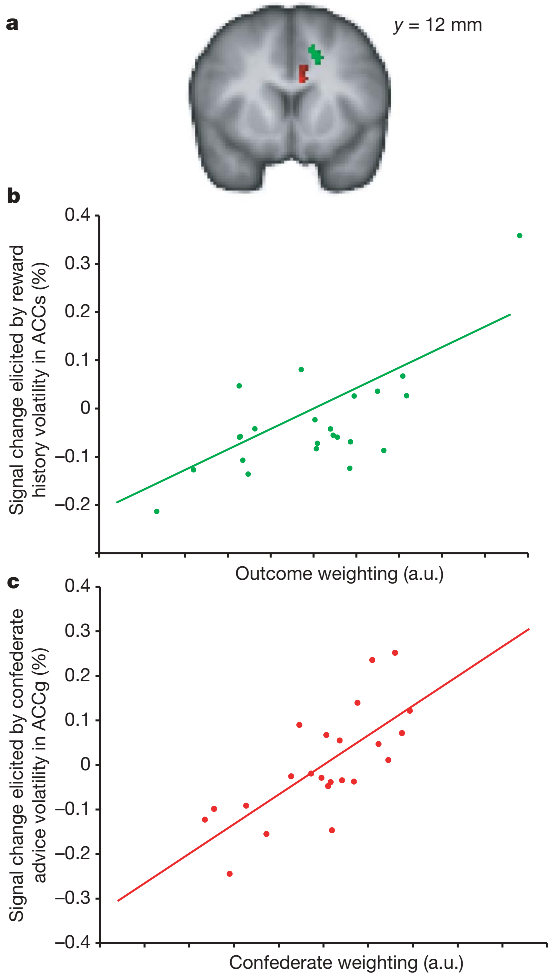

Figure 3 | Agency-specific learning rates dissociate in the ACC. **a**, Regions where the BOLD correlates of reward (green) and confederate (red) volatility predict the influence that each source of information has on subject behaviour ($Z > 3.1$, $P < 0.05$ cluster-corrected for cingulate cortex). **b**, Subjects with high BOLD signal changes in response to reward volatility in the ACCs are guided strongly by reward history information (maximum $Z = 3.7$, correlation $R = 0.7163$, $P < 0.0001$). **c**, Subjects with high BOLD signal changes in response to confederate advice volatility in the ACCg are guided strongly by social information (maximum $Z = 4.1$, correlation $R = 0.7252$, $P < 0.0001$). See Supplementary Information.

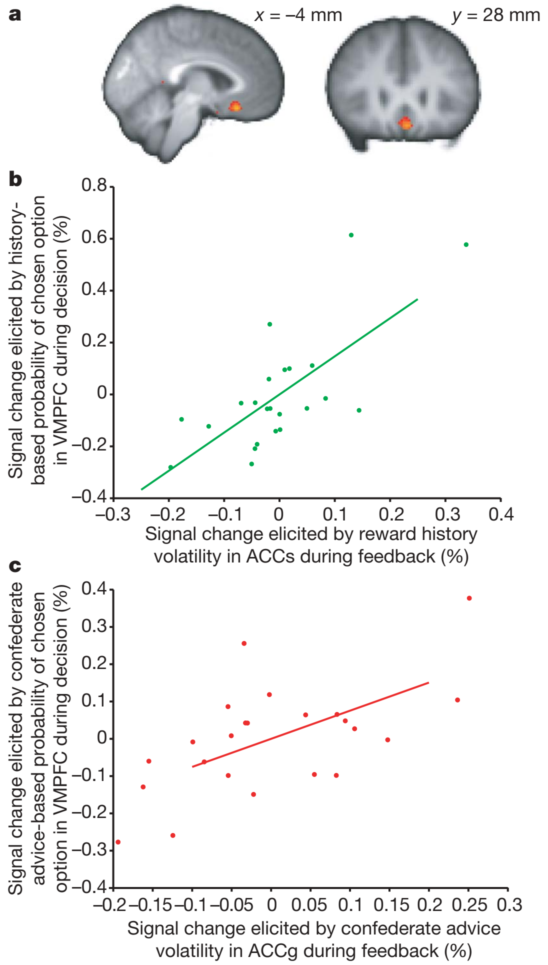

Figure 4 | Combination of expected value of chosen option in VMPFC. **a**, Activation for the combination (mean contrast) of experience-based probability during the Cue and Suggest phases, and advice-based probability during the Suggest phase (threshold set at $Z > 3.1$, $P < 0.005$ cluster-corrected for VMPFC). These phases represent the times at which subjects had these probabilities available to them (see Supplementary Fig. 4). **b**, Correlation between the effect of outcome-based probability in VMPFC during the decision and the effect of outcome volatility in ACCs during the Monitor phase ($R = 0.7113$, $P < 0.0002$). **c**, Correlation between the effect of confederate-based probability in VMPFC during the decision and the effect of confederate volatility in ACCs during the Monitor phase ($R = 0.6119$, $P < 0.002$). See Supplementary Information.

of information from social and experiential sources. However, the extent to which the VMPFC signal reflects each source of information during a decision is predicted by the extent to which the ACCs and ACCg modulate their activity at the point when information is learnt. If, as is suggested, the VMPFC response codes the expected value of a decision, then the ACCs response to each new outcome predicts the extent that this outcome will determine future valuation of an action; the ACCg response predicts the extent to which this outcome will determine future valuation of an individual.

METHODS SUMMARY

Short description of task (Fig. 1a). Subjects performed a decision-making task while undergoing fMRI, repeatedly choosing between blue and green rectangles, each of which had a different reward magnitude available on each trial. The chance of the rewarded colour being blue or green depended on the recent outcome history. Before the experiment, subjects were introduced to a confederate. At each trial, the confederate would choose between supplying the subject with the correct or incorrect option, unaware of the number of points available. The subject's goal was to maximize the number of points gained during the experiment. In contrast, the confederate's goal was to ensure that the eventual score would lie within one of two pre-defined ranges, known to the confederate but not the subject. The confederate might therefore reasonably give consistently helpful or unhelpful advice, but this advice might change as the game progressed (Supplementary Information). During the experiment, the confederate was replaced by a computer that gave correct advice on a prescribed set of trials. Subjects knew that the trial outcomes were determined by an inanimate computer program, but believed that the social advice came from an animate agent's decision.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 September; accepted 14 October 2008.

1. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425,** 785–791 (2003).
2. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, 1982).
3. Delgado, M. R., Frank, R. H. & Phelps, E. A. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neurosci.* **8,** 1611–1618 (2005).
4. King-Casas, B. *et al.* Getting to know you: reputation and trust in a two-person economic exchange. *Science* **308,** 78–83 (2005).
5. Rilling, J. *et al.* A neural basis for social cooperation. *Neuron* **35,** 395–405 (2002).

6. Gallagher, H. L., Jack, A. I., Roepstorff, A. & Frith, C. D. Imaging the intentional stance in a competitive game. *Neuroimage* **16,** 814–821 (2002).
7. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
8. Behrens, T. E., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. Learning the value of information in an uncertain world. *Nature Neurosci.* **10,** 1214–1221 (2007).
9. Courville, A. C., Daw, N. D. & Touretzky, D. S. Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.* **10,** 294–300 (2006).
10. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nature Neurosci.* **3** (Suppl.), 1218–1223 (2000).
11. O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304,** 452–454 (2004).
12. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
13. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412,** 43–48 (2001).
14. Matsumoto, M., Matsumoto, K., Abe, H. & Tanaka, K. Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neurosci.* **10,** 647–656 (2007).
15. Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neurosci.* **7,** 887–893 (2004).
16. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441,** 876–879 (2006).
17. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47,** 129–141 (2005).
18. Haruno, M. & Kawato, M. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J. Neurophysiol.* **95,** 948–959 (2006).
19. D'Ardenne, K., McClure, S. M., Nystrom, L. E. & Cohen, J. D. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* **319,** 1264–1267 (2008).
20. Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J. & Rushworth, M. F. Optimal decision making and the anterior cingulate cortex. *Nature Neurosci.* **9,** 940–947 (2006).
21. Deaner, R. O., Khera, A. V. & Platt, M. L. Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* **15,** 543–548 (2005).
22. Shepherd, S. V., Deaner, R. O. & Platt, M. L. Social status gates social attention in monkeys. *Curr. Biol.* **16,** R119–R120 (2006).
23. Rudebeck, P. H., Buckley, M. J., Walton, M. E. & Rushworth, M. F. A role for the macaque anterior cingulate gyrus in social valuation. *Science* **313,** 1310–1312 (2006).
24. O'Doherty, J. P. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* **14,** 769–776 (2004).
25. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nature Neurosci.* **10,** 1625–1633 (2007).
26. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nature Rev. Neurosci.* **7,** 268–277 (2006).
27. Allison, T., Puce, A. & McCarthy, G. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* **4,** 267–278 (2000).
28. Castelli, F., Frith, C., Happe, F. & Frith, U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* **125,** 1839–1849 (2002).
29. Van Hoesen, G. W., Morecraft, R. J. & Vogt, B. A. in *Neurobiology of Cingulate Cortex and Limbic Thalamus* (eds Vogt, B. A. & Gabriel, M.) (Birkhäuser, 1993).
30. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27,** 9984–9988 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments We would like to acknowledge funding from the UK MRC (T.E.J.B., M.F.S.R.), the Wellcome Trust (L.T.H.) and the UK EPSRC (M.W.W.). We thank S. Knight for helping with data acquisition, and K. Watkins for help with figure preparation.

Author contributions All four authors contributed to generating the hypothesis and designing the experiment. Where specific roles can be assigned, L.T.H. collected the data, T.E.J.B. and L.T.H. analysed the data, T.E.J.B. and M.W.W. built the model, and M.F.S.R. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.E.J.B. (behrens@fmrib.ox.ac.uk).

METHODS

Detailed analysis of the task, the learning model, the behavioural analysis, the data acquisition and pre-processing, and several further results and discussion can be found in the Supplementary Information. Here, we describe aspects of the fMRI modelling that may be relevant to the interpretation of our results. Further technical details can also be found in the Supplementary Information.

fMRI single-subject modelling. We performed two fMRI GLM analyses using FMRIB's Software library (FSL, ref. 31). The first looked for learning-related activity (Figs 2, 3 and Supplementary Fig. 3), the second for decision-related activity (Fig. 4 and Supplementary Fig. 4). In each case a general linear model was fit in pre-whitened data space (to account for autocorrelation in the fMRI residuals)[32]. Regressors were convolved and filtered according to FSL defaults (see Supplementary Information).

The following regressors (plus their temporal derivatives) were included in the time series model (learning-related activity): four regressors defining the different times during the task (see Fig. 1 and Supplementary Information), namely Cue, Suggest, Interval, Monitor; four regressors defining key learning parameters when the outcomes are presented (see Supplementary Information), namely (Monitor × Reward history volatility), (Monitor × Confederate volatility), (Monitor × Reward prediction error), (Monitor × Confederate prediction error).

The following regressors (plus their temporal derivatives) were included in the time series model (decision-related activity): four regressors defining the different times during the task (see Fig. 1 and Supplementary Information), namely Cue, Suggest, Interval, Monitor; seven regressors defining key decision parameters at the times when they were available during the decision (see Supplementary Information), namely (Cue × Experience-based probability), (Suggest × Experienced-based probability), (Suggest × Confederate-based probability), (Cue × Chosen reward magnitude), (Suggest × Chosen reward magnitude), (Cue × Unchosen reward magnitude), (Suggest × Unchosen reward magnitude). Note that probabilities were log-transformed such that their linear combination in the GLM would approximate the optimal combination for behaviour (see Supplementary Information). Figure 4a was generated using the mean ([1 1 1]) contrast of all probability-related regressors.

fMRI group modelling. fMRI group analyses were carried out using a GLM with three regressors: a group mean, the weight for reward history information based on each subject's behaviour (see Supplementary Information), and the weight for confederate information based on each subject's behaviour (see Supplementary Information).

fMRI region of interest analyses (Fig. 2). The following processing steps are illustrated schematically in Supplementary Fig. 2 and described in more detail in the Supplementary Information. Individual subject data were taken from regions of interest defined by the group clusters. Data from each trial were up-sampled and re-aligned to points in the trial corresponding to the onset of the four trial stages. Data were Z-normalized across trials at each time point in the trial. We then performed two general linear models across trials for both reward and confederate prediction errors. This allowed us (1) to test at which points in the trial the data correlated with the prediction of reward, or the prediction of confederate fidelity, and (2) to test at which points after the outcome the data correlated with the trial outcome, or actual confederate fidelity. A prediction error signal should comprise three parts. (1) A positive correlation with the prediction after the decision; (2) a positive correlation with the trial outcome at the time of this outcome; (3) a negative correlation with the prediction at the time of the outcome (as a prediction error is defined as the outcome minus the prediction).

We witnessed all three parts of the confederate prediction error as deflections in BOLD correlations at the relevant times. However, owing to the nature of the haemodynamic response, it is difficult to test significance from just these deflections. We therefore fit a haemodynamic model to these correlation profiles in each subject (see Supplementary Information). The key test was whether the time course of correlations with the prediction could be accounted for by a positive haemodynamic impulse at the time of the decision and a negative haemodynamic impulse at the time of the outcome; and whether the time course of correlations with the outcome could be accounted for by a positive haemodynamic impulse at the time of the outcome. By fitting the haemodynamic model we were able to measure three parameter estimates for each of these three haemodynamic impulses in each subject, and perform random-effects t-tests to measure statistical significance of each.

31. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23** (Suppl. 1), S208–S219 (2004).

32. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage* **14,** 1370–1386 (2001).