

Neural mechanisms of observational learning

Christopher J. Burke^{a,1,2}, Philippe N. Tobler^a, Michelle Baddeley^b, and Wolfram Schultz^a

^aDepartment of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom; and ^bFaculty of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom

Edited by Leslie G. Ungerleider, National Institute of Mental Health, Bethesda, MD, and approved July 6, 2010 (received for review March 12, 2010)

Individuals can learn by interacting with the environment and experiencing a difference between predicted and obtained outcomes (prediction error). However, many species also learn by observing the actions and outcomes of others. In contrast to individual learning, observational learning cannot be based on directly experienced outcome prediction errors. Accordingly, the behavioral and neural mechanisms of learning through observation remain elusive. Here we propose that human observational learning can be explained by two previously uncharacterized forms of prediction error, observational action prediction errors (the actual minus the predicted choice of others) and observational outcome prediction errors (the actual minus predicted outcome received by others). In a functional MRI experiment, we found that brain activity in the dorsolateral prefrontal cortex and the ventromedial prefrontal cortex respectively corresponded to these two distinct observational learning signals.

prediction error | reward | vicarious learning | dorsolateral prefrontal cortex | ventromedial prefrontal cortex

In uncertain and changing environments, flexible control of actions has individual and evolutionary advantages by allowing goal-directed and adaptive behavior. Flexible action control requires an understanding of how actions bring about rewarding or punishing outcomes. Through instrumental conditioning, individuals can use previous outcomes to modify future actions (1–4). However, individuals learn not only from their own actions and outcomes but also from those that are observed. One of the most illustrative examples of observational learning happens in Antarctica, where flocks of Adelle penguins often congregate at the water's edge to enter the sea and feed on krill. However, the main predator of the penguins, the leopard seal, is often lurking out of sight beneath the waves, making it a risky prospect to be the first one to take the plunge. As this waiting game develops, one of the animals often becomes so hungry that it jumps, and if no seal appears the rest of the group will all follow suit. The following penguins make a decision after observing the action and outcome of the first (5). This ability to learn from observed actions and outcomes is a pervasive feature of many species and can be absolutely crucial when the stakes are high. For example, predator avoidance techniques or the eating of a novel food item are better learned from another's experience rather than putting oneself at risk with trial-and-error learning. Although we know a fair amount about the neural mechanisms of individuals learning about their own actions and outcomes (6), almost nothing is known about the brain processes involved when individuals learn from observed actions and outcomes (7). This lack of knowledge is all the more surprising given that observational learning is such a wide-ranging phenomenon.

In this study, 21 participants engaged in a novel observational learning task based on a simple two-armed bandit problem (Fig. 1A) while being scanned. On a given trial, participants chose one of two abstract fractal stimuli to gain a stochastic reward or to avoid a stochastic punishment. One stimulus consistently delivered a good outcome (reward or absence of punishment) 80% of the time and a bad outcome (absence of reward or punishment) 20% of the time. The other stimulus consistently had opposite outcome contingencies (20% good outcome and 80% bad outcome). The participants' task was to learn to choose the better of the two stimuli. However, before the participants made their own choice, they were able to observe the behavior of a confederate player who

was given the same stimuli to choose from. As such, participants had access to two sources of information to help them learn which stimulus was the best; they could observe the other player's actions and outcomes and also learn from their own reinforcement given their own action. In an individual learning baseline condition, no information about the confederate's actions and outcomes was available. Thus, participants could learn the task only through their own outcomes and actions (individual trial and error). In an impoverished observational learning condition, the actions but not the outcomes of the confederate player were available. Finally, in a full observational learning condition, the amount of information shown to participants was maximized by displaying both the actions and outcomes of the confederate player. Thus, in both the impoverished and the full observational learning conditions participants could learn not only from individual but also from external sources.

We hypothesized that the ability of humans to learn from observed actions and outcomes could be explained using well-established reinforcement learning models. In particular, we explored the possibility that two previously uncharacterized forms of prediction error underlie the ability to learn from others and that these parameters are represented in a similar manner to those seen in the brain during individual learning. Prediction errors associated with an event can be defined as the difference between the prediction of an event and the actual occurrence of an event. Thus, on experiencing an event repeatedly, prediction errors can be computed and then used by an organism to increase the accuracy of their prediction that the same event will occur in the future (learning). We adapted a standard action-value learning algorithm from computational learning theory to encompass learning from both individual experience and increasing amounts of externally observed information. By using a model-based functional MRI (fMRI) approach (8) we were able to test whether activity in key learning-related structures of the human brain correlated with the prediction errors hypothesized by the models.

Results

Participants learned to choose the better stimulus more often with increasing amounts of observable information (ANOVA, $P < 0.001$, Fig. 1B and D; on the individual subject level, this relation was significant in 18 of the 21 subjects). When the actions and the outcomes of the confederate were observable, participants chose the correct stimulus at a significantly higher rate compared with when only confederate actions were observable (ANOVA, $P < 0.01$) and during individual learning ($P < 0.001$). In addition, when only confederate actions were observable, the participants

Author contributions: C.J.B., P.N.T., M.B., and W.S. designed research; C.J.B. and P.N.T. performed research; C.J.B. analyzed data; and C.J.B., P.N.T., M.B., and W.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹Present address: Institute for Empirical Research in Economics, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland.

²To whom correspondence should be addressed. E-mail: burke@iew.uzh.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1003111071/-DCSupplemental.

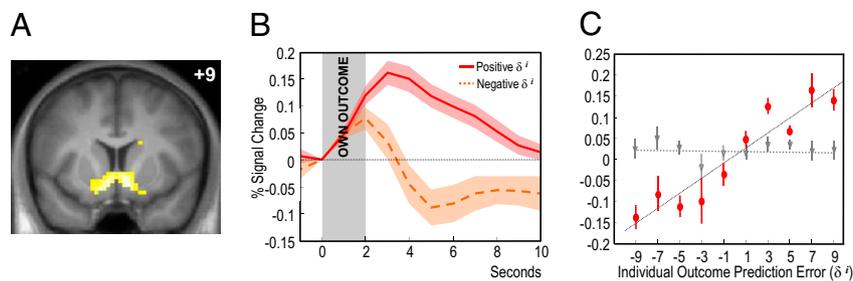


Fig. 2. Activity in the ventral striatum correlating with individual outcome prediction error. (A) Coronal section showing significant voxels correlating with individual outcome prediction errors ($P < 0.05$, whole-brain correction). (B) Time course of activity in the ventral striatum binned according to sign of individual outcome prediction error. (C) Linear regression of activity in the ventral striatum against individual outcome prediction errors as expected by the model (red circular markers, $P < 0.001$, $R^2 = 0.888$). To demonstrate that social and nonsocial learning are integrated at the neural level, the gray triangular markers show the regression with expected outcome prediction errors when social information is removed from the model on social learning trials ($P = 0.917$, $R^2 = 0.001$).

choice by the confederate is unexpected would drive attention and, by extension, learning in the present context through an unsigned prediction error. This unsigned action prediction error is multiplied by an imitation factor (analogous to the learning rate in individual reinforcement learning) to update the probability of the participant choosing that stimulus the next time they see it. The participant then refines the values associated with the stimuli by experiencing individual outcomes (and therefore individual outcome prediction errors) when making decisions between the respective stimuli. The values associated with a particular stimulus pair are used to update the individual probabilities of choosing each stimulus via a simple choice rule (softmax). These probabilities are also the predicted probabilities of the choice of the other player. In such a way, combining imitative responses with directly received individual reinforcement allows the observing individual to increase the speed at which a correct strategy is acquired compared with purely individual learning (13).

At the time of the choice of the other player in these imitative learning trials, we found a highly significant activation in the dorsolateral prefrontal cortex (DLPFC) [48, 30, 27, $Z = 4.52$, $P < 0.05$ small volume correction (SVC)] corresponding to the action prediction error proposed in our imitation learning model (Fig. 3A). Time-courses from the DLPFC showed that this previously uncharacterized observational learning signal differentiated between small and large action prediction errors maximally between 3 and 6 s after the presentation of the choice of the other player (Fig. 3B). The region showed a monotonic increase in activity with increasing levels of the action prediction error expected by our imitative reinforcement learning model (Fig. 3C). This signal can be thought of as the degree of unpredictability of the actual choice of the other player relative to the predicted choice probability. In other words, if the scanned participant predicts that the confederate will choose stimulus *A* with a high probability, the action prediction error would be small if the confederate subsequently makes that choice. On early trials, such a signal effectively biases

the participant to imitate the other player's action (i.e., select the same stimulus). Note that the imitative prediction error signal is specific for stimulus choice and cannot be explained through simple motor/direction imitation because the position of the stimuli on the other player's and participant's screens were varied randomly.

The dorsolateral prefrontal cortex was the only region that showed an action prediction error at the time of choice of the other player. Notably, there were no significant voxels in the ventral striatum that correlated with the expected action prediction error signal as predicted by our model, even at the low statistical threshold of $P < 0.01$, uncorrected. Upon occurrence of individual outcomes however, a prediction error signal was again observed in the ventral striatum (9, 6, -9, $Z = 4.91$, $P < 0.023$ whole-brain correction), lending weight to the idea that a combination of observed action prediction error and individual outcome prediction errors drive learning when only actions are observable. When the model was run as if observable information was ignored by the participants (e.g., testing the hypothesis that participants only learned from their own rewards), the regression of ventral striatum activity against expected individual outcome prediction error became insignificant ($R^2 = 0.001$, $P = 0.917$), indicating an integration of social and individual learning at the neural level (Fig. 2C).

In the fully observable condition, we further increased the level of information available to the scanned participants and allowed them to observe not only confederate actions but also confederate outcomes, in addition to their own outcomes. In a similar manner to the imitative learning model, we suggest that learning in this type of situation is governed by a two-stage updating process, driven by observational as well as individual outcome prediction errors. However, as outcomes have incentive value, we used signed rather than unsigned prediction error signals in both the observed and the individual case. In the model, observational outcome prediction errors arise at the outcome of the other player and the observer updates their outcome expectations in a manner similar to individual learning. This necessitates the processing of outcomes that

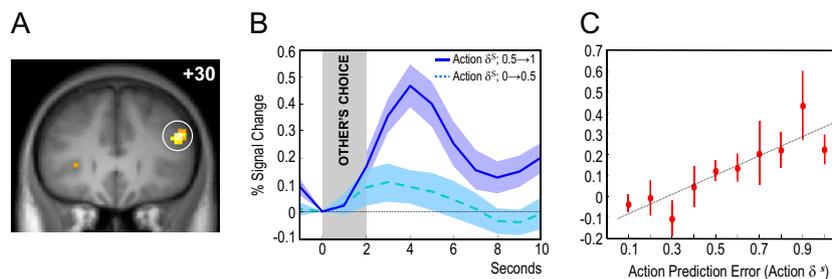


Fig. 3. Activity in the DLPFC correlating with observational action prediction error. (A) Coronal section showing significant voxels correlating with action prediction errors ($P < 0.05$, SVC for frontal lobe). (B) Time course of activity in DLPFC at the time of the other player's choice, binned according to magnitude of action prediction error. (C) Linear regression of activity in DLPFC with action prediction errors expected by the model (red circular markers, $P < 0.001$, $R^2 = 0.773$).

are not directly experienced by the observer. Similarly to individual prediction errors, observational outcome prediction errors serve to update the observer's probabilities of choosing a particular stimulus next time they see it. Finally, learning is refined further by the ensuing individual prediction error. We found a region of ventromedial prefrontal cortex (VMPFC) that significantly correlated with an observational outcome prediction error signal (peak at $-6, 30, -18, Z = 3.12, P < 0.05$ SVC) at the time of the outcome of the other player (Fig. 4A). This signal showed maximal differentiation between positive and negative observational prediction errors between 4 and 6 s after the presentation of the outcome of the other player. Also at the time of the outcome of the other player, we observed the inverse pattern of activation in the ventral striatum (i.e., increased activation with decreasing magnitude of observational prediction error; peak at $-12, 12, -3, Z = 4.07, P < 0.003$, SVC) (Fig. 4D–F). In other words, the striatum was activated by observed outcomes worse than predicted and deactivated by outcomes better than predicted. Conversely, during the participants' outcomes in the full observational condition, ventral striatum activity reflected the usual expected individual outcome prediction error, with (de)activations to outcomes (worse) better than predicted ($Z = 5.49, P < 0.001$ whole-brain corrected). Taken together, these data suggest that the VMPFC processes the degree to which the actual outcome of the other player was unpredicted relative to the individual's prediction, whereas the ventral striatum emits standard and inverse outcome prediction error signals in experience and observation respectively.

Discussion

In the present study, we show that human participants are able to process and learn from observed outcomes and actions. By incorporating two previously uncharacterized forms of prediction error into simple reinforcement learning algorithms, we have been able to investigate the possible neural mechanisms of two aspects of observational learning.

In the first instance, learning from observing actions can be explained in terms of an action prediction error, coded in the dorsolateral prefrontal cortex, corresponding to the discrepancy between the expected and actual choice of an observed individual. Upon experiencing the outcomes of their own actions, learners can then combine what they have learned based on action prediction errors with individual learning based on outcome prediction errors. Through such a combination, the learner acquires a better prediction of others' future actions. The application of a simple reinforcement learning rule to such behavior is tractable as it does not require the observer to remember the previous sequence of choices and outcomes (14). Small action prediction errors allow the learning participant to reconfirm that they are engaging in the correct strategy without observing the outcome of the other player. The region of DLPFC where activity correlates most strongly with the expected action prediction error signal has previously been shown to respond to other types of prediction error (15), to conflict (16), and to trials that violate an expectancy that was learned from previous trials (17). The present finding also fits well with previous research implicating the DLPFC in action selection (18). In particular, DLPFC activity increases with increasing uncertainty regarding which action to select (19). In the present task, uncertainty about which action to select may have been particularly prevalent when subjects were learning which action to select through observation, an interpretation that could be partially supported by the finding of increases in imitative responses in the action only observation condition.

In the second instance, the VMPFC processed observational outcome prediction errors. This learning signal applies to observed outcomes that could never have been obtained by the individual (as opposed to actually or potentially experienced outcomes) and drives the observational (or vicarious) learning seen in our two-stage bandit task. In a distinct form of learning (fictive or counterfactual learning), agents learn from the outcomes that they could have received, had they chosen differently. Fictive reward signals (rewards that could have been, but were not directly received) have

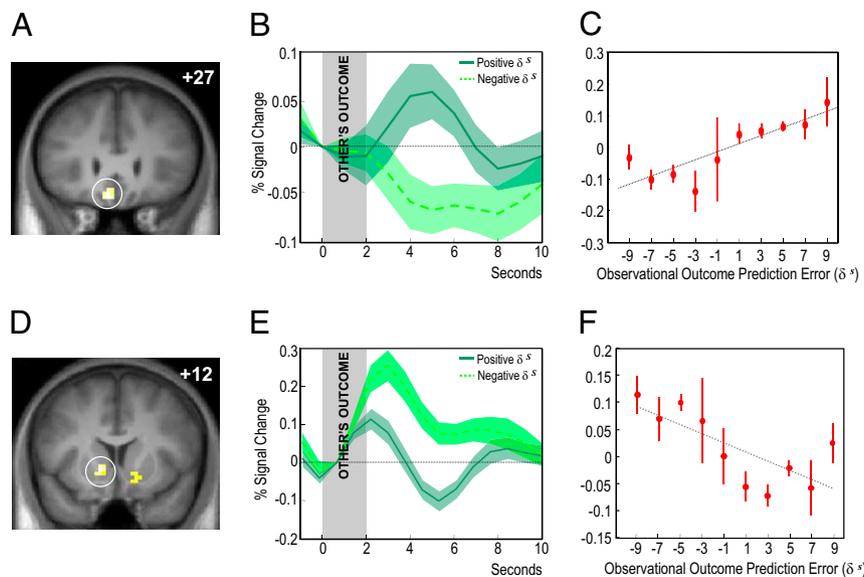


Fig. 4. Activity in VMPFC and ventral striatum corresponding to observational outcome prediction errors during the outcome of the other player. (A) Coronal section showing significant voxels correlating with observational outcome prediction errors in VMPFC [$P < 0.05$, SVC for 30 mm around coordinates of reported peak in O'Doherty et al. (38)]. (B) Time course of activity in VMPFC during the outcome of the other player, binned according to sign of observed outcome prediction error. (C) Linear regression of activity in VMPFC at the time of the other player's outcome with expected observational outcome prediction errors (red circular markers, $P < 0.002$, $R^2 = 0.719$). (D) Coronal section showing significant voxels correlating with inverse observational outcome prediction errors in ventral striatum ($P < 0.003$, SVC). (E) Time course of activity in ventral striatum during the outcome of the other player, binned according to sign of observational prediction errors. (F) Linear regression of activity in ventral striatum at the time of the other player's outcome with expected observational prediction errors (red circular markers, $P < 0.03$, $R^2 = -0.552$).

been previously documented in the anterior cingulate cortex (20) and the inferior frontal gyrus (21). These regions contrast with those presently observed, further corroborating the fundamental differences between fictive and the presently studied vicarious rewards. Fictive rewards are those that the individual could have received (but did not receive), whereas vicarious rewards are those that another agent received but were never receivable by the observing individual. Outside the laboratory, vicarious rewards are usually spatially separated from the individual, whereas fictive rewards tend to be temporally separated (what could have been, had the individual acted differently). The observational outcome prediction errors in the present study may be the fundamental learning signal driving vicarious learning, which has previously been unexplored from combined neuroscientific and reinforcement learning perspectives.

In addition to the fundamental difference in the present study's findings on observational learning and those on fictive learning, a major development is the inclusion of two learning algorithms that can explain observational learning using the standard reinforcement learning framework (these models are depicted graphically in Figs. S3 and S4, with parameter estimates shown in Fig. S5). These models should be of considerable interest to psychologists working on observational learning in animals and behavioral ecologists. To our knowledge, standard reinforcement learning methods have not previously been used to explain this phenomenon.

Of particular relevance to our results, the VMPFC has been previously implicated in processing reward expectations based on diverse sources of information, and activity in this region may represent the value of just-chosen options (22). Our data add to these previous findings by showing an extension to outcomes resulting from options just chosen by others. Taken together, the present results suggest a neural substrate of vicarious reward learning that differs from that of fictive reward learning.

Interestingly, it appears that the ventral striatum, an area that has been frequently associated with the processing of prediction errors in individual learning (23–26), was involved in processing prediction errors related to actually experienced, individual reward in our task in a conventional sense (i.e., with increasing activity to positive individual reward prediction errors) but showed the inverse coding pattern for observational prediction errors. Although our task was not presented to participants as a game situation (and the behavior of the confederate in no way affected the possibility of participants receiving reward), this inverse reward prediction error coding for the confederate's outcomes is supported by previous research highlighting the role the ventral striatum plays in competitive social situations. However, the following interpretations must be considered with reservation, especially because of the lack of nonsocial control trials in our task. For instance, the ventral striatum has been shown to be activated when a competitor is punished or receives less money than oneself (27). This raises a number of interesting questions for future research on the role of the ventral striatum in learning from others. For example, do positive reward prediction errors when viewing another person lose drive observational learning, or is it simply rewarding to view the misfortunes of others? Recent research suggests that the perceived similarity in the personalities of the participant and confederate modulates ventral striatum activity when observing a confederate succeed in a nonlearning game show situation (28). During action-only learning, the individual outcome prediction error signal emitted by the ventral striatum can serve not only to optimize one's own outcome-oriented choice behavior but (in combination with information on what others did in the past) can also refine predictions of others' choice behavior when they are in the same choice situation.

We found observational outcome and individual outcome prediction errors as well as prediction errors related to the actions of others. They are all computed according to the same principle of comparing what is predicted and what actually occurs. Neverthe-

less, we show that different types of prediction error signals are coded in distributed areas of the brain. Taken together, our findings lend weight to the idea that the computation of prediction errors may be ubiquitous throughout the brain (29–31). This predictive coding framework has been shown to be present during learning unrelated to reward (32, 33) and for highly cognitive concepts such as learning whether or not to trust someone (34). Indeed, an interesting extension in future research on observational learning in humans would be to investigate the role "social appraisal" plays during learning from others. For example, participants may change the degree to which they learn from others depending on prior information regarding the observed individual, and outcome-related activity during iterated trust games has been shown to be modulated by perceptions of the partner's moral character (35).

These findings show the utility of a model-based approach in the analysis of brain activity during learning (36). The specific mechanisms of observational learning in this experiment are consistent with the idea that it is evolutionary efficient for general learning mechanisms to be conserved across individual and vicarious domains. Previous research on observational learning in humans has focused on the learning of complex motor sequences that may recruit different mechanisms, with many studies postulating an important role for the mirror system (37). However, the role that the mirror system plays in observationally acquired stimulus-reward mappings remains unexplored. The prefrontal regions investigated in this experiment may play a more important role in inferring the goal-directed actions of others to create a more accurate picture of the local reward environment. This idea is borne out in recent research suggesting that VMPFC and other prefrontal regions (such as more dorsal regions of prefrontal cortex and the inferior frontal gyrus) are involved in a mirror system that allows us to understand the intentions of others. This would allow the individual to extract more reward from the environment than would normally be possible relying only on individual learning with no possibility to observe the mistakes or success of others.

Materials and Methods

Participants. A total of 23 right-handed healthy participants were recruited through advertisements on the University of Cambridge campus and on a local community website. Two participants were excluded for excessive head motion in the scanner. Of those scanned and retained, 11 were female and the mean age of participants was 25.3 y (range 18–38 y). Before the experiment, participants were matched with a confederate volunteer who was previously unknown to them (*SI Text*).

Behavioral Task—Sequence of Events Each trial of the task started with a variable intertrial interval (ITI) with the fixation cross-presented on the side of the screen dedicated to the other player (Fig. 1A). This marked the beginning of the "observation stage," which would later in the trial be followed by the "action stage." During the observation stage, the photo of the other player was displayed on their half of the screen at all times. The side of the screen assigned to the other player was kept fixed throughout the experiment, but counterbalanced across participants to control for visual laterality confounds. The ITI varied according to a truncated Poisson distribution of 2–11 s. After fixation, two abstract fractal stimuli were displayed on the other player's screen for 2 s. When the fixation cross was circled, the player in the scanner was told that the other player must choose between the two stimuli within a time window of 1 s. The participant in the scanner also had to press the third button of the response pad during this 1-s window in order for the trial to progress. This button press requirement provided a basic motor control for motor requirements in the action stage and ensured attentiveness of participants during the observation stage. Depending on the trial type, participants were then shown the other player's choice by means of a white rectangle appearing around the chosen stimulus for 1 s. On trial types in which the action of the other player was unobservable, both stimuli were surrounded by a rectangle so the actual choice could not be perceived by the participant. The outcome of the other player was then displayed for 2 s. On trial types in which no outcome was shown, a scrambled image with the same number of pixels as unscrambled outcomes was displayed.

After the end of observation stage, the fixation cross switched to the participant's side of the screen and another ITI (with the same parameters as previously mentioned) began. This marked the beginning of the action stage of a trial. During the action stage, the photograph of the player inside the scanner was displayed on their half of the screen at all times. The participant in the scanner was presented with the same stimuli as previously shown to the other player. The left/right positions of the stimuli were randomly varied at the observation and action stages of each trial to control for visual and motor confounds. When the fixation cross was circled, participants chose the left or right stimuli by pressing the index or middle finger button on the response box respectively. Stimulus presentation and timing was implemented using Cogent Graphics (Wellcome Department of Imaging Neuroscience, London, United Kingdom) and Matlab 7 (Mathworks).

Trial Types and Task Sessions. Three trial types were used to investigate the mechanisms of observational learning. The trial types differed according to the amount of observable information available to the participant in the scanner (Table S1). Participants underwent six sessions of ≈ 10 min each in the scanner. Three of these sessions were "gain" sessions and three were "loss" sessions. During gain sessions, the possible outcomes were 10 and 0 points, and during loss sessions, the possible outcomes were 0 and 10 points. Gain and loss sessions alternated, and participants were instructed before each session started as to what type it would be. In each session, participants learned to discriminate between three pairs of stimuli. Within a session, one of three stimulus pairs was used for each of the three trial types (coinciding with the different levels of information regarding the confederate's behavior).

The trial types experienced by the participant were randomly interleaved during a session, although the same two stimuli were used for each trial type. For example, all "full observational" learning trials in a single session would use the same stimuli and "individual" learning trials would use a different pair. On a given trial, one of the two stimuli presented produced a "good" outcome (i.e., 10 in a gain session and 0 in a loss session) with probability 0.8 and a bad outcome (i.e., 0 in a gain session and -10 in a loss session) with a probability of 0.2. For the other stimulus, the contingencies were reversed. Thus, participants had to learn to choose the correct stimulus over the course of a number of trials. New stimuli were used at the start of every session, and participants experienced each trial type 10 times per session (giving 60 trials per trial type over the course of the experiment). Time courses in the three regions of interest over the course of a full trial can be seen in Fig. S6. Lists of significant activation clusters in all three conditions can be seen in Table S2.

Image Analysis. We conducted a standard event-related analysis using SPM5 (Functional Imaging Laboratory, University College London, available at www.fil.ion.ucl.ac.uk/spm/software/spm5). Prediction errors generated by the computational models were used as parametric modulation of outcome regressors (SI Materials and Methods).

ACKNOWLEDGMENTS. We thank Anthony Dickinson and Matthijs van der Meer for helpful discussions. This work was supported by the Wellcome Trust and the Leverhulme Trust.

- Balleine BW, Dickinson A (1998) Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–419.
- Thorndike EL (1911) *Animal Intelligence: Experimental Studies* (Macmillan, New York).
- Mackintosh NJ (1983) *Conditioning and Associative Learning* (Oxford University Press, New York).
- Skinner B (1938) *The Behavior of Organisms: An experimental analysis* (Appleton-Century, Oxford).
- Chamley C (2003) *Rational Herds: Economic Models of Social Learning* (Cambridge University Press, New York).
- O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr Opin Neurobiol* 14:769–776.
- Subiaul F, Cantlon JF, Holloway RL, Terrace HS (2004) Cognitive imitation in rhesus macaques. *Science* 305:407–410.
- O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci* 1104:35–53.
- Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: Expectation and prediction. *Psychol Rev* 88:135–170.
- Widrow G, Hoff M (1960) Adaptive Switching Circuits. *Institute of Radio Engineers Western Electronic Show and Convention* (Convention Record, Institute of Radio Engineers, New York), pp 96–104.
- Kettner RE, et al. (1997) Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement. *J Neurophysiol* 77:2115–2130.
- Pearce JM, Hall G (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87:532–552.
- Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci USA* 105:6741–6746.
- Fel'tovich N (2000) Reinforcement-based vs. beliefs-based learning in experimental asymmetric-information games. *Econometrica* 68:605–641.
- Fletcher PC, et al. (2001) Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat Neurosci* 4:1043–1048.
- Menon V, Adelman NE, White CD, Glover GH, Reiss AL (2001) Error-related brain activation during a Go/NoGo response inhibition task. *Hum Brain Mapp* 12:131–143.
- Casey BJ, et al. (2000) Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proc Natl Acad Sci USA* 97:8728–8733.
- Rowe JB, Toni I, Josephs O, Frackowiak RSJ, Passingham RE (2000) The prefrontal cortex: Response selection or maintenance within working memory? *Science* 288:1656–1660.
- Frith CD (2000) The role of dorsolateral prefrontal cortex in the selection of action as revealed by functional imaging. *Control of Cognitive Processes. Attention and Performance XVI/11*, eds Monsell S, Driver J (MIT Press, Cambridge, MA), pp 549–565.
- Hayden BY, Pearson JM, Platt ML (2009) Fictive reward signals in the anterior cingulate cortex. *Science* 324:948–950.
- Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci USA* 104:9493–9498.
- Rushworth MF, Mars RB, Summerfield C (2009) General mechanisms for making decisions? *Curr Opin Neurobiol* 19:75–83.
- Bray S, O'Doherty J (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol* 97:3036–3045.
- O'Doherty JP, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
- Pagnoni G, Zink CF, Montague PR, Berns GS (2002) Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci* 5:97–98.
- Rodriguez PF, Aron AR, Poldrack RA (2006) Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Hum Brain Mapp* 27:306–313.
- Fliessbach K, et al. (2007) Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318:1305–1308.
- Mobbs D, et al. (2009) A key role for similarity in vicarious reward. *Science* 324:900.
- Bar M (2007) The proactive brain: Using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289.
- Friston K, Kilner J, Harrison L (2006) A free energy principle for the brain. *J Physiol Paris* 100:70–87.
- Schultz W, Dickinson A (2000) Neuronal coding of prediction errors. *Annu Rev Neurosci* 23:473–500.
- den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. *Cereb Cortex* 19:1175–1185.
- Summerfield C, Koehlin E (2008) A neural representation of prior information during perceptual inference. *Neuron* 59:336–347.
- Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. *Nature* 456:245–249.
- Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* 8:1611–1618.
- Behrens TEJ, Hunt LT, Rushworth MF (2009) The computation of social behavior. *Science* 324:1160–1164.
- Catmur C, Walsh V, Heyes C (2007) Sensorimotor learning configures the human mirror system. *Curr Biol* 17:1527–1531.
- O'Doherty J, et al. (2003) Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147–155.