Neural mechanisms of observational learning

Christopher J. Burke[a,1,2], Philippe N. Tobler[a], Michelle Baddeley[b], and Wolfram Schultz[a]

[a]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom; and [b]Faculty 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 Adelie 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. 1*A*) 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. 1 *B* 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.

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

[2]To 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.1003111107/-/DCSupplemental.

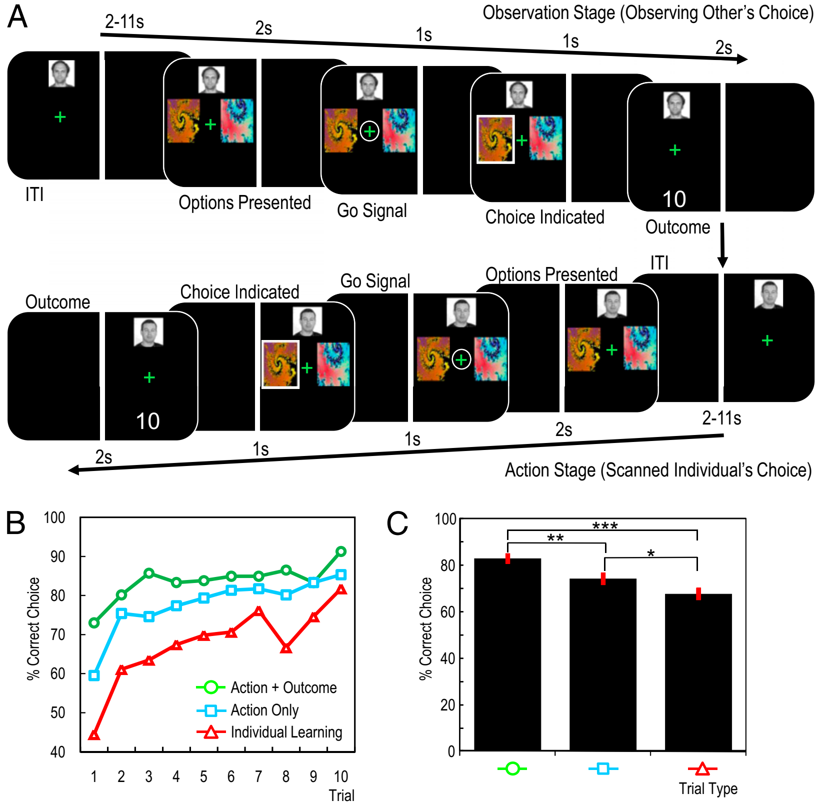

Fig. 1. Experimental design and behavioral results. (*A*) After a variable ITI, participants were first given the opportunity to observe the confederate player being presented with two abstract fractal stimuli to choose from. After another variable ITI, participants were then presented with the same stimuli, and the trial proceeded in the same manner. When the fixation cross was circled, participants made their choice using the index finger (for left stimulus) and middle finger (for right stimulus) on the response pad. (*B*) The proportion of correct choices increased with increasing amounts of social information. (*C*) There was a monotonic increase in the proportion of correct choices as a function of the observability of the other player's behavior and outcomes. Learning from the actions and outcomes of the other player resulted in significantly more correct choices than action only observable and individual learning conditions.

still chose the correct stimulus on a significantly higher proportion of trials compared with the individual baseline ($P < 0.05$). Thus, more observable information led to more correct choice.

As would be expected, more observable information also led to higher earnings. In general, there was a monotonic increase in the amount of reward received (in gain sessions) and a decrease in the amount of punishment received (in loss sessions) with increasing observable information. Although the confederate's behavior did not differ across conditions (Fig. S1), Significantly more reward and less punishment was received by participants when they observed the confederate's actions and outcomes in comparison with the individual baseline ($P < 0.001$ and $P < 0.001$, ANOVA; Fig. S2). The behavioral data demonstrate that participants were able to use observed outcomes and actions to improve their performance in the task.

Next we investigated whether participants relied on imitation in the conditions in which actions or actions and outcomes of the confederate were observable. Imitation occurred in the present context when participants chose the same stimulus rather than the same motion direction as confederates (because the positions of stimuli were randomized across confederate and participants periods within a trial). Participants showed more imitative responses when only the actions of confederates were observable (on 66.4% of all trials) compared with when both actions and outcomes were observable (on 41.1% of all trials; two-tailed *t* test, $P < 0.001$). This finding, together with the finding that participants made more correct choices and earned more money when both actions and outcomes were observable, suggests that they adaptively used available information in the different conditions.

To relate individual learning to brain activity when participants had no access to external information, we fitted a standard action-value learning algorithm onto individual behavior and obtained individual learning rates. Based on these individual learning rates, we computed expected prediction errors for each participant in each trial. These values correspond to the actual outcome participants received minus the outcome they expected from choosing a given stimulus. Finally, we entered the expected outcome prediction error values into a parametric regression analysis and located brain activations correlating with expected individual prediction errors. In concordance with previous studies on individual action–outcome learning, brain activations reflected prediction error signals in the ventral striatum (peak at 9, 9, −12, $Z = 5.96$, $P < 0.05$, whole-brain corrected) (Fig. 2*A*).

In a second condition, we increased the level of information available to participants and allowed them to observe the actions, but not the outcomes, of the confederate player. Contrary to the individual learning baseline condition, in this environment participants can infer the outcome received by the other player and observationally learn from the other player's actions. Subsequently, they can combine this inference with individual learning. Such a process can be thought of as a two-stage learning process; first, observing the action of another player biases the observer to imitate that action (at least on early trials), and second, the outcome of the observer's own action refines the values associated with each stimulus. Learning from the actions but not the outcomes of others is usually modeled using sophisticated Bayesian updating strategies, but we hypothesized it could also be explained using slight adaptations to the standard action-value model used in individual learning, by the inclusion of a novel form of prediction error. Imitative responses can be modeled using an "action prediction error" which corresponds to the difference between the expected probability of observing a choice and the actual choice made, analogous to the error terms in the delta learning rule (9, 10) and some forms of motor learning (11). Contrary to an incentive outcome, a simple motor action does not have either rewarding or punishing value in itself. Accordingly, the best participants could do was to pay more attention to unexpected action choices by the confederate as opposed to expected action choices. Attentional learning models differ from standard reinforcement learning models in that they use an unsigned rather than a signed prediction error (12). Accordingly, the degree to which an action

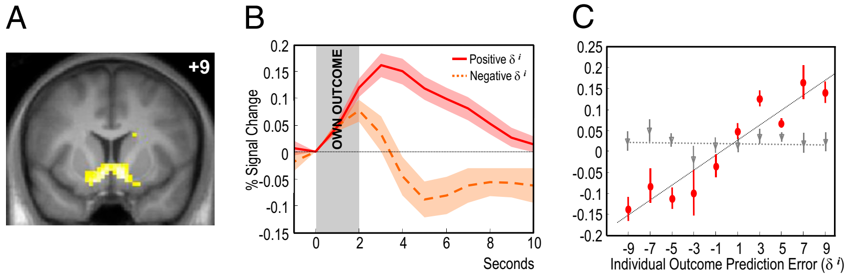

Fig. 2. Activity in the ventral striatum correlating with individual outcome prediction error. (*A*) Coronal section showing significant voxels correlating with individual outcome prediction errors ($P < 0.05$, whole-brain correction). (*B*) Time course of activity in the ventral striatum binned according 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. 3*A*). 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. 3*B*). The region showed a monotonic increase in activity with increasing levels of the action prediction error expected by our imitative reinforcement learning model (Fig. 3*C*). 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. 2*C*).

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

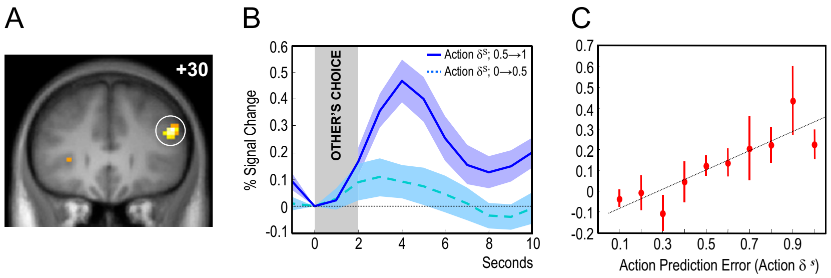

Fig. 3. Activity in the DLPFC correlating with observational action prediction error. (*A*) Coronal section showing significant voxels correlating with action prediction errors ($P < 0.05$, SVC for frontal lobe). (*B*) Time course of activity in DLPFC at the time of the other player's choice, binned 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. 4*A*). 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. 4 *D*–*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

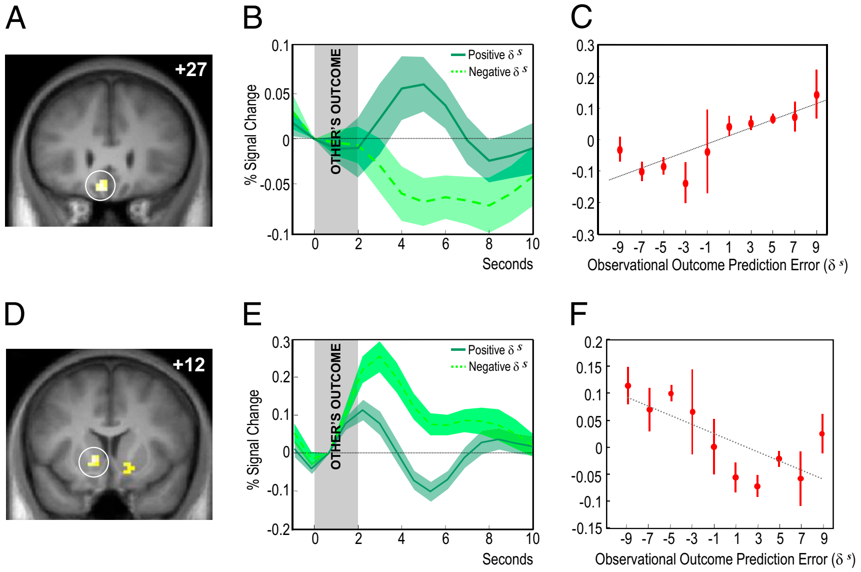

Fig. 4. Activity in VMPFC and ventral striatum corresponding to observational outcome prediction errors during the outcome of the other player. (*A*) Coronal section showing significant voxels correlating with observational outcome prediction errors in VMPFC [$P < 0.05$, SVC for 30 mm around coordinates 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 errror. (*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. Nevertheless, 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. 1*A*). 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 1s. 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.

1. Balleine BW, Dickinson A (1998) Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–419.
2. Thorndike EL (1911) *Animal Intelligence: Experimental Studies* (Macmillan, New York).
3. Mackintosh NJ (1983) *Conditioning and Associative Learning* (Oxford University Press, New York).
4. Skinner B (1938) *The Behavior of Organisms: An experimental analysis* (Appleton-Century, Oxford).
5. Chamley C (2003) *Rational Herds: Economic Models of Social Learning* (Cambridge University Press, New York).
6. O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr Opin Neurobiol* 14:769–776.
7. Subiaul F, Cantlon JF, Holloway RL, Terrace HS (2004) Cognitive imitation in rhesus macaques. *Science* 305:407–410.
8. 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.
9. Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: Expectation and prediction. *Psychol Rev* 88:135–170.
10. 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.
11. 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.
12. 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.
13. 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.
14. Feltovich N (2000) Reinforcement-based vs. beliefs-based learning in experimental asymmetric-information games. *Econometrica* 68:605–641.
15. 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.
16. Menon V, Adleman 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.
17. 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.
18. 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.
19. 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 XVI11*, eds Monsell S, Driver J (MIT Press, Cambridge, MA), pp 549–565.
20. Hayden BY, Pearson JM, Platt ML (2009) Fictive reward signals in the anterior cingulate cortex. *Science* 324:948–950.
21. 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.
22. Rushworth MF, Mars RB, Summerfield C (2009) General mechanisms for making decisions? *Curr Opin Neurobiol* 19:75–83.
23. Bray S, O'Doherty J (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol* 97:3036–3045.
24. O'Doherty JP, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
25. 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.
26. Rodriguez PF, Aron AR, Poldrack RA (2006) Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Hum Brain Mapp* 27: 306–313.
27. Fliessbach K, et al. (2007) Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318:1305–1308.
28. Mobbs D, et al. (2009) A key role for similarity in vicarious reward. *Science* 324:900.
29. Bar M (2007) The proactive brain: Using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289.
30. Friston K, Kilner J, Harrison L (2006) A free energy principle for the brain. *J Physiol Paris* 100:70–87.
31. Schultz W, Dickinson A (2000) Neuronal coding of prediction errors. *Annu Rev Neurosci* 23:473–500.
32. 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.
33. Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. *Neuron* 59:336–347.
34. Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. *Nature* 456:245–249.
35. 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.
36. Behrens TEJ, Hunt LT, Rushworth MFS (2009) The computation of social behavior. *Science* 324:1160–1164.
37. Catmur C, Walsh V, Heyes C (2007) Sensorimotor learning configures the human mirror system. *Curr Biol* 17:1527–1531.
38. O'Doherty J, et al. (2003) Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147–155.

Supporting Information

Burke et al. 10.1073/pnas.1003111107

SI Materials and Methods

Participants. All participants were fluent speakers of English and had normal or corrected-to-normal vision in the scanner. Participants were preassessed to exclude previous histories of neurological or psychiatric illness. All participants gave informed consent, and the Local Research Ethics Committee of the Cambridgeshire Health Authority approved the study. To minimize error trials during scanning, participants learned the timings and sequence of task events (for 20 training trials per condition with stimuli not used in the experiment) no more than 7 d before scanning. During the training period, black and white portrait photographs were taken of each participant against a plain white background at a fixed distance of 2 m. The images were cropped to 100 × 100 pixels and adjusted to have equal luminance. During training, participants were instructed that they would be taking part in a social experiment with two players. They were instructed that they would be able to observe the behavior of another player but that the other player would not be able to observe them. When participants arrived at the scanner, an experimental confederate arrived a little later. The participants were gender matched to confederates. Confederates and participants sat together in the waiting area of the MR facility and went through the same procedures with regards to filling in forms, reading task instructions and being checked for metals. After these preliminary procedures, one research team member led the confederate into another room, where another computer was present. Another member of the research team led the true participant into the scanner. After scanning, the exit of the confederate from the facility was timed to coincide with the debriefing of the true participant (who was sat in the waiting area). The confederates never actually performed the task (except to familiarize themselves with the experiment), and the behavior of what the true participant believed to be the other player was controlled by a computer and kept constant across participants. There was very little difference in the performance of the computer over the trial types (Fig. S1), indicating that the differential participant performance according to trial type was a function of the amount of social information available.

Participant Payment. Participants were paid according to the total points accumulated during all sessions of the task, which were converted to British pounds sterling at a rate of 30 points to the pound. In accordance with local payment protocol, participants also received 20 pounds for participating regardless of task performance. The average participant payment was 52 pounds.

Data Acquisition. Scanning took place at the Medical Research Council's Cognition and Brain Sciences Unit (MRC-CBU), Cambridge, United Kingdom. The task was projected on a display, which participants viewed through a mirror fitted on top of the head coil. We acquired gradient echo T2*-weighted echo-planar images (EPIs) with blood-oxygen-level–dependent (BOLD) contrast on a Siemens Trio 3 Tesla scanner (slices/volume, 33; repetition time, 2 s). Depending on performance of participants, 280–350 volumes were collected in each session of the experiment, together with five "dummy" volumes at the start and end of each scanning session. Scan onset times varied randomly relative to stimulus onset times.

A T1-weighted MP-RAGE structural image was also acquired for each participant. Signal dropout in basal frontal and medial temporal structures resulting from susceptibility artifact was reduced by using a tilted plane of acquisition (30° to the anterior commissure-posterior commissure line, rostral > caudal). Imaging parameters were the following: echo time, 50 ms; field of view, 192 mm. The in-plane resolution was 3 × 3 mm, with a slice thickness of 2 mm and an interslice gap of 1 mm. High-resolution T1-weighted structural scans were coregistered to their mean EPIs and averaged together to permit anatomical localization of the functional activations at the group level.

Image Analysis. We used a standard rapid-event–related fMRI approach in which evoked hemodynamic responses to each event type are estimated separately by convolving a canonical hemodynamic response function with the onsets for each event and regressing these against the measured fMRI signal (1, 2). This approach makes use of the fact that the hemodynamic response function summates in an approximately linear manner over time (3). By presenting trials in strictly random order and using randomly varying intertribal intervals, it is possible to separate out fMRI responses to rapidly presented events without waiting for the hemodynamic response to reach baseline after each single trial (1, 2).

Statistical parametric mapping (SPM5; Functional Imaging Laboratory, University College London, available at www.fil.ion.ucl.ac.uk/spm/software/spm5) served to spatially realign functional data, normalize them to a standard EPI template and smooth them using an isometric Gaussian kernel with a full-width at half-maximum of 8 mm. Onsets of stimuli and outcomes were modeled as separate delta functions and convolved with a canonical hemodynamic response function. Participant-specific movement parameters were modeled as covariates of no interest. Linear contrasts of regression coefficients were computed at the individual subject level and then taken to group-level *t* tests.

Computational Models. We adapted a basic Q learning algorithm that has been previously shown to account for instrumental choice in probabilistic reward-learning tasks (4, 5). Generally, for a given binary choice between two stimuli (A and B), the standard Q learning model estimates the expected value of choosing A or B. Whenever an outcome is observed for choosing a particular stimulus at time *t*, a prediction error (δ) (corresponding to the realized minus the expected outcome) is computed. The Q value associated with that stimulus is updated accordingly by multiplying the prediction error by the learning rate (α). At the start of a session, the Q values associated with each stimulus were set to zero. If, for example, on the first trial the subject chose stimulus A and received an outcome (r) of 10 points, the prediction error δ would be given by $\delta(t) = r(t) - Qa(t)$. The value of stimulus A would then be updated according to $Qa(t+1) = Qa(t)+\alpha^*\delta(t)$. The probability of the model subsequently selecting a stimulus was determined using the softmax function (6). The softmax function computes a probability of selecting a particular stimulus from a pair according to the ratio of the Q values associated with each stimulus and parameter β (the inverse temperature, which captures the degree of variability in choices). The softmax function has been shown to provide a good approximation of binary choice in previous experiments (4).

Full observational learning. During full observational learning (i.e., when the action and outcome of the other player was observable), the standard Q learning algorithm was modified by incorporating a two-stage update process per trial (Fig. S3). The first update occurs during the "observation stage" and the second during the "action stage" (Fig. 1*A*). As such, the first update (after the observation stage) occurs at *t*+0.5, halfway through the trial. Upon observing an outcome received by the other player, the scanned participant is assumed to experience an observational reward prediction error (δ^S), according to the outcome received by the confederate (r^S) minus the Q value associated with that stimulus. This trial-by-trial observational outcome prediction error was entered as a parametric modulator at the onset of the other player's outcome. The scanned participant is able to learn from the reinforcement received by the other player by multiplying the social reward prediction error (δ^S) by the observational learning rate (α^S), capturing the degree to which participants are able to learn from outcomes that are not directly experienced. This update results in an observationally-updated Q value at time $t+0.5$ (denoted by Q^S in Fig. S1 for display purposes).

At the choice of the participant (during the action stage), the probabilities of choosing a particular stimulus are modeled using the softmax function, taking the observationally updated Q values as arguments. Upon receipt of individual outcome (r), an individual reward prediction error (δ^i) is computed by subtracting the previous observationally updated Q value (Q^S) from the individual outcome (r). These trial-by-trial values were entered into the general linear model as a parametric modulator at the onset of the participant's outcome. The Q values are then updated according to the standard algorithm by multiplying δ^i with the individual learning rate (α^i).

Action imitation learning. When only the action of the other player is observable, learning can be modeled with the incorporation of action prediction errors with the standard Q learning algorithm (Fig. S4). The protocol follows the two-stage update procedure outlined previously. At the start of the observation stage, the participant in the scanner has Q values associated with each stimulus on the screen, and therefore has some probability of choosing each stimulus according to the softmax function. For example, at the start of a session when no learning has occurred the ratio of the two stimulus values gives a 0.5 probability of choosing a particular stimulus. When the confederate goes on to make a choice, the action prediction error (δ^{action}) is given by the actual choice minus the probability of choice associated with that stimulus from softmax. Because the actual choice is always 1 or 0, action prediction errors are always in the positive domain. For example, if at the start of the session the probability associated with choosing stimulus A is 0.5 (as no learning has occurred and the Q values for the stimuli are equal), the action prediction error would be [action (a) = 1] – [probability of choosing (a) = 0.5] = 0.5. The degree to which the participant incorporates this information in driving his subsequent choice behavior is controlled by an imitation factor (κ) analogous to the learning rate in the standard algorithm. Therefore, the probability of the participant subsequently choosing A is $P(a)_{(t+0.5)} = P(a)_{(t)} + \kappa^* \delta^{action}$. Conversely, the probability of choosing B is simply $1 - P(a)$. Although no actual outcome has been observed, after a number of trials the ratio of the stimulus values is inferred. This ratio of action probabilities drives the scanned participant's choice in an analogous fashion to softmax. When the scanned participant subsequently chooses and receives an outcome from a particular stimulus, the participant computes an individual reward prediction error and update the Q values according to the standard algorithm, thereby refining value estimations.

Individual learning. On individual learning trials, the choice and outcome of the confederate player were not observable. On such trials, at the time of the confederate's choice during the observation stage both stimuli were surrounded by white rectangles, making it impossible for the scanned participant to determine which was chosen. At the time of the confederate's outcome, a scrambled image was displayed at the same location. On these trials, participants were required to learn from only their received reinforcements, and the standard Q-learning algorithm was used to model this.

To generate the regressors for the novel prediction error signals, the free parameters in each model (α^i, α^S, κ, and β) were adjusted to maximize the likelihood of observing each participant's choices given the respective model according to $L = \prod^{N}_{n=1}\prod^{T}_{t=1}P_{(choice,n,t)}$, where N is the number of participants; T is the number of trials per participant, and $P_{(choice,n,t)}$ is the likelihood of choice made by participant n at trial t given the model. MATLAB (Mathworks) was used to find the parameters maximizing the likelihood L, with values of the parameters searched in increments of 0.01 from 0 to 1, the results of which can be seen in Fig. S5. In such a manner, individual, session-specific regressors for theoretical social, individual, and action prediction errors were generated and subsequently tested for covariation with brain signals.

Behavioral Results. As noted in the main text, there was a significant increase in participants' performances with increasing amounts of observable information [ANOVA, F(2,21) = 11.305, $P < 0.001$]. For example, participants chose the correct stimulus at a significantly higher rate when they were able to observe the actions and outcomes of the confederate compared with when only confederate actions were observable ($P < 0.01$). In turn, the participants did significantly better when actions were observable compared with learning without any observable information ($P < 0.05$).

When the data were split according to gain and loss sessions (Fig. S2), the monotonic increase in performance with observable information was preserved ($P < 0.001$ in gain sessions, $P < 0.02$ in loss sessions). In both gain and loss scenarios, participants chose the correct stimulus at a significantly higher rate and received significantly more points in the fully observable condition compared with the individual learning baseline. However, in gain sessions, there was a significant difference between fully observable and action-only conditions for both correct choices and reward received ($P < 0.001$ and $P < 0.001$ respectively) but not between action-only and individual conditions. In loss session, this pattern was reversed (no significant differences between fully observable and action-only conditions for both performance and points received). However, a two-way ANOVA with sessions (gain/loss) and trial type (fully observable, action-only, and individual learning conditions) as factors failed to show a significant interaction ($P = 0.42$). In summary, it appears that participants were able to increase their performance by learning from the actions of the confederate more efficiently in loss sessions as opposed to gain sessions. Previous research has suggested that learning from positive and negative reinforcement may be mediated by opposing neural systems (7) and individual differences in the effectiveness of learning from rewards and punishments have been documented (8). One possibility underlying the differences we observed could be that learning through observing the actions of others is more effective in situations where an avoidance response is necessary, numerous examples of which exist in the literature (9, 10). Indeed, research in foraging theory predicts that observational learning should proceed more readily in resource-poor environments or for the learning of predator avoidance mechanisms (11).

1. Dale AM, Buckner RL (1997) Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* 5:329–340.
2. Josephs O, Henson RN (1999) Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philos Trans R Soc Lond B Biol Sci* 354: 1215–1228.
3. Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221.
4. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442: 1042–1045.
5. Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
6. Luce RD (1986) *Response Times: Their Role in Inferring Elementary Mental Organisation* (Oxford University Press, New York).

7. Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. *Neural Netw* 15:603–616.
8. Frank MJ, Woroch BS, Curran T (2005) Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* 47:495–501.
9. Mineka S, Davidson M, Cook M, Keir R (1984) Observational conditioning of snake fear in rhesus monkeys. *J Abnorm Psychol* 93:355–372.
10. Olsson A, Phelps EA (2004) Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychol Sci* 15:822–828.
11. Laland KN (2004) Social learning strategies. *Learn Behav* 32:4–14.

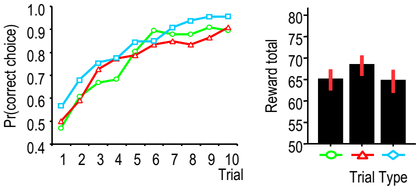

Fig. S1. Computer-controlled confederate's behavioral performance was constant across trial type.

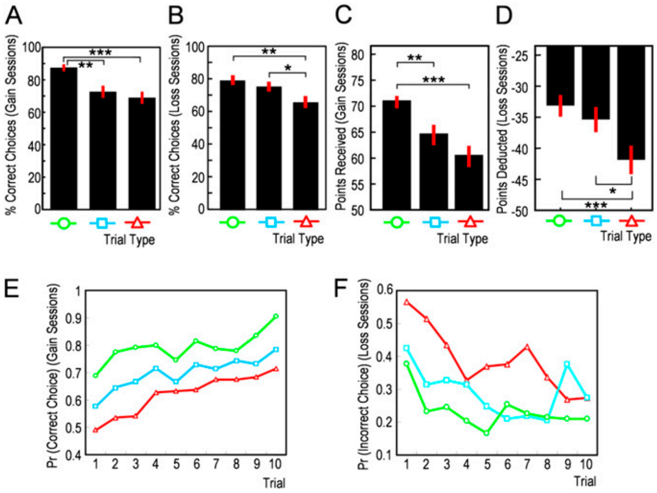

Fig. S2. Percentage of correct choices in gain (*A*) and loss (*B*) sessions and points scored by participants in gain (*C*) and loss sessions (*D*), separated according to trial type. (*E*) Probabilities of correct choices on a trial-by-trial basis in gain sessions. (*F*) Probabilities of correct choices on a trial-by-trial basis in loss sessions.

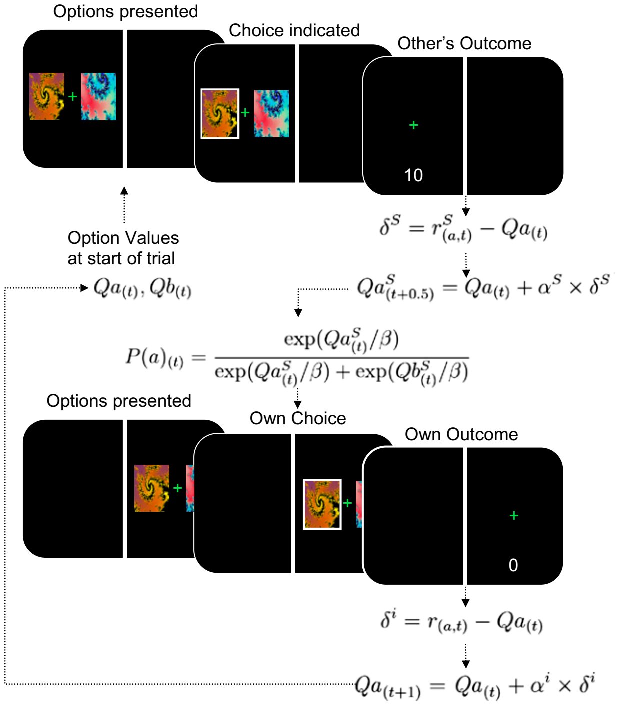
Options presented
Choice indicated
Other's Outcome
10
$\delta^S = r^S_{(a,t)} - Qa_{(t)}$
Option Values
at start of trial
$Qa_{(t)}, Qb_{(t)}$
$Qa^S_{(t+0.5)} = Qa_{(t)} + \alpha^S \times \delta^S$
$P(a)_{(t)} = \frac{\exp(Qa^S_{(t)}/\beta)}{\exp(Qa^S_{(t)}/\beta) + \exp(Qb^S_{(t)}/\beta)}$
Options presented
Own Choice
Own Outcome
0
$\delta^i = r_{(a,t)} - Qa_{(t)}$
$Qa_{(t+1)} = Qa_{(t)} + \alpha^i \times \delta^i$

Fig. S3. Schematic diagram illustrating the learning model for when all social information is available (i.e., when the action and outcome of the confederate player is observable). The two-stage learning process is illustrated as if stimulus A has been chosen by both confederate and participant, denoted in the lowercase and subscript text. In this particular example, the confederate received 10 points and the participant 0 points for their choices.

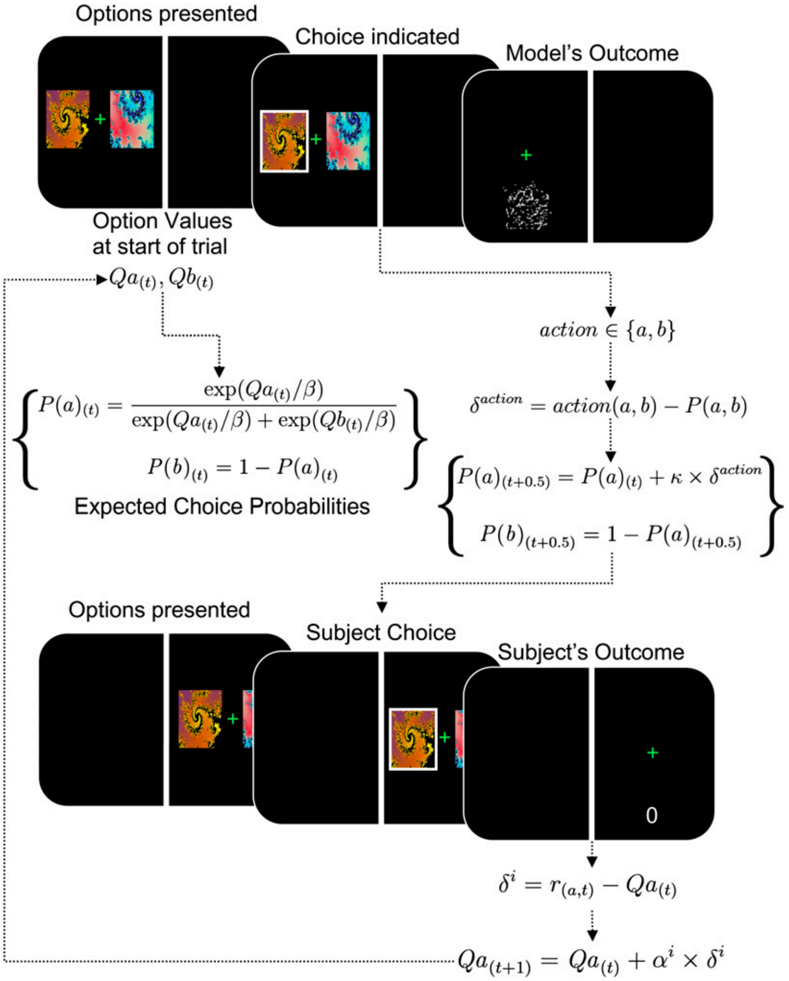
Options presented
Choice indicated
Model's Outcome
Option Values at start of trial
$Qa_{(t)}, Qb_{(t)}$
$action \in \{a, b\}$
$P(a)_{(t)} = \frac{\exp(Qa_{(t)}/\beta)}{\exp(Qa_{(t)}/\beta) + \exp(Qb_{(t)}/\beta)}$
$P(b)_{(t)} = 1 - P(a)_{(t)}$
Expected Choice Probabilities
$\delta^{action} = action(a, b) - P(a, b)$
$P(a)_{(t+0.5)} = P(a)_{(t)} + \kappa \times \delta^{action}$
$P(b)_{(t+0.5)} = 1 - P(a)_{(t+0.5)}$
Options presented
Subject Choice
Subject's Outcome
0
$\delta^{i} = r_{(a,t)} - Qa_{(t)}$
$Qa_{(t+1)} = Qa_{(t)} + \alpha^{i} \times \delta^{i}$

Fig. S4. Schematic illustrating the learning model for when only the actions of the other player are observable. In this particular example, both confederate and participant chose stimulus A, denoted in the lowercase and subscript text.

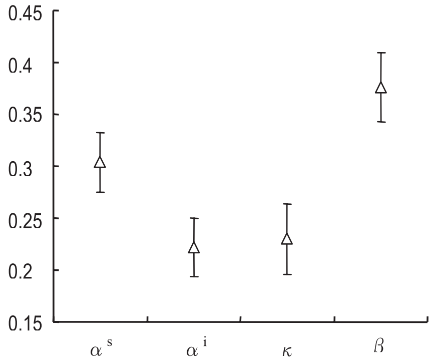

Fig. S5. Free parameter values that maximized the likelihood of observing participants' behavioral data given the social learning models.

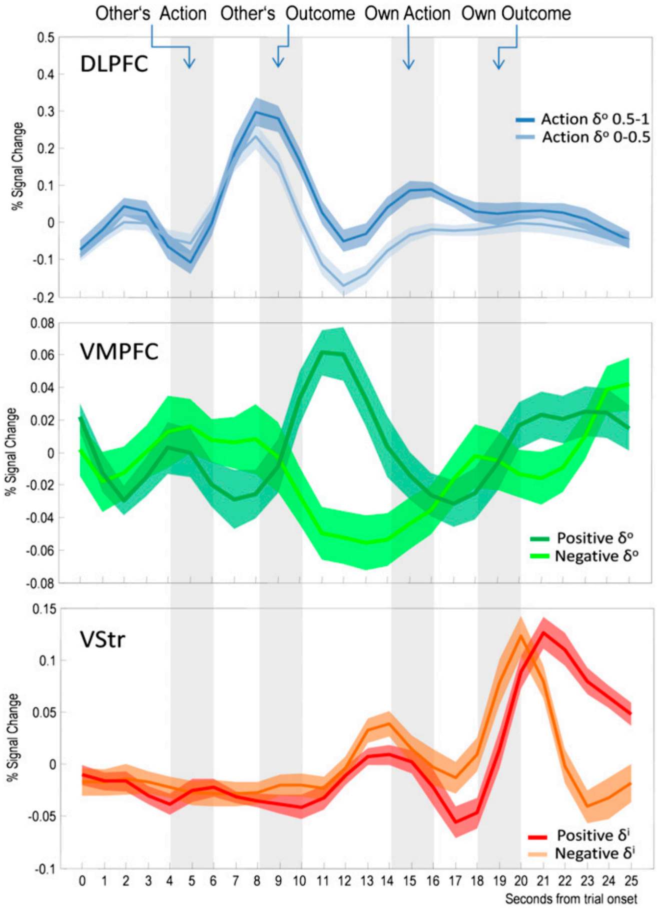

Fig. S6. Time course in the three regions of interest (DLPFC, VMPFC, and ventral striatum) over the course of a full trial.

Table S1. Trial types summarized in terms of the availability of social information to the participant in the scanner

Learning type	Other's action	Other's outcome
Full observational learning	Observable	Observable
Action imitation learning	Observable	Hidden
Nonobservational learning	Hidden	Hidden

Colored shapes are the same as those used in the main text.

Table S2. Locations of significant activation clusters for the action learning, action + outcome learning, and individual learning conditions in a whole-brain analysis

Cluster location	MNI X (mm)	MNI Y (mm)	MNI Z (mm)	No. of voxels	Peak z score
Action learning condition					
R inferior temporal gyrus	48	−48	−15	74	4.25
R middle occipital gyrus	39	−72	21	27	3.98
R inferior occipital gyrus	39	−72	−15	16	3.52
Full observational learning					
L insula	−42	−6	3	12	2.93
R medial orbitofrontal cortex	12	45	−12	3	2.56
L medial orbitofrontal cortex	−3	39	−15	2	2.48
Individual learning					
L precuneus	−27	−60	21	45	5.91
R superior occipital gyrus	30	−78	0	43	5.80
L angular gyrus	−48	−72	36	35	5.60
L middle temporal gyrus	−54	−3	−24	44	5.60
L cerebellum	−21	−57	−45	13	5.54
R cerebellum	42	−66	−42	149	5.53
R middle temporal gyrus	60	0	−15	30	5.30
L paracentral Lobule	−18	−9	66	28	4.13
L rolandic opercularis	−57	15	12	20	3.90

Montreal Neurological Institute (MNI) coordinates denote the peak of each cluster. Activations at $P < 0.001$ uncorrected with an extent threshold of 10 voxels are listed. However, for the full observational learning condition, no other activations were observed at the usual threshold of $P < 0.001$ with an extent threshold of 10 voxels. As such, activations at $P < 0.01$ uncorrected with an extent threshold of 0 voxels are listed for this contrast. R, right; L, left.