

# Behavioral and Neural Properties of Social Reinforcement Learning

Rebecca M. Jones,<sup>1</sup> Leah H. Somerville,<sup>1</sup> Erika J. Ruberry,<sup>1</sup> Victoria Libby,<sup>1</sup> Gary Glover,<sup>2</sup> Henning U. Voss,<sup>3</sup> Douglas J. Ballarín,<sup>4</sup> and B. J. Casey<sup>1</sup>

<sup>1</sup>Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, New York, New York 10021, <sup>2</sup>Department of Psychology, New York University, New York, New York 10003, <sup>3</sup>Stanford Center for Imaging, Department of Radiology, Stanford University, Stanford, California 94305, and <sup>4</sup>Citigroup Biomedical Imaging Center, Department of Radiology, Weill Cornell Medical College, New York, New York, 10065

Social learning is critical for engaging in complex interactions with other individuals. Learning from positive social exchanges, such as acceptance from peers, may be similar to basic reinforcement learning. We formally test this hypothesis by developing a novel paradigm that is based on work in nonhuman primates and human imaging studies of reinforcement learning. The probability of receiving positive social reinforcement from three distinct peers was parametrically manipulated while brain activity was recorded in healthy adults using event-related functional magnetic resonance imaging. Over the course of the experiment, participants responded more quickly to peers who provided more frequent positive social reinforcement, and rated them as more likeable. Modeling trial-by-trial learning showed ventral striatum and orbital frontal cortex activity correlated positively with forming expectations about receiving social reinforcement. Rostral anterior cingulate cortex activity tracked positively with modulations of expected value of the cues (peers). Together, the findings across three levels of analysis—social preferences, response latencies, and modeling neural responses—are consistent with reinforcement learning theory and nonhuman primate electrophysiological studies of reward. This work highlights the fundamental influence of acceptance by one's peers in altering subsequent behavior.

## Introduction

Successfully navigating our social environment depends on learning from positive and negative encounters with others and shaping future behavior toward those individuals. Psychologists have proposed that positive social exchanges are fundamentally rewarding for humans (Bandura and Walters, 1963; Baumeister and Leary, 1995; Steinberg, 2008), suggesting that learning from social interactions may draw on basic reinforcement learning mechanisms. The present study was designed to test this hypothesis by building on reinforcement learning studies in nonhuman primates and human imaging studies (Schultz et al., 1997; Fiorillo et al., 2003; McClure et al., 2003; D'Ardenne et al., 2008).

Reinforcement learning from primary (e.g., food) and secondary (e.g., money) reinforcers has been shown to engage specific neural circuitry. In its simplest form, it is explained by the classic Rescorla–Wagner model (Rescorla and Wagner, 1972). According to this model, learning to associate arbitrary cues with

greatest probability of providing social acceptance to the participant. We applied a simple Rescorla–Wagner rule in behavioral and functional imaging analyses to target the neural bases of these behavioral changes, hypothesizing that the ventral striatum and OFC would code prediction error signals (Schultz et al., 1997; O’Doherty, 2007). Thus, the current study elucidates neurobiological mechanisms for key learning processes during social exchanges that shape behavior through positive interactions.

## Materials and Methods

**P** . Forty-six adults (aged 18–28 years; 22 females) participated in the experiment. Thirty-six completed the task during functional magnetic resonance imaging (fMRI) (aged 18–28 years; all right-handed; 19 females). Three individuals in the fMRI group were excluded due to insufficient number of correct trials in any condition ( $n = 2$ ; 1 male) or noncompliance with the task ( $n = 1$ , male). Participants reported no history of neurological and/or psychiatric disorders in a standard screening or on the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) (First et al., 2007) and imaging participants reported no contraindications for an MRI. Two participants did not complete the SCID due to time constraints. All participants provided written consent approved by the Institutional Review Board at Weill Cornell Medical College and were debriefed and compensated following their participation.

**E** . The experiment was conducted during two separate sessions. The first session introduced the cover story, leading participants to believe they would receive actual social feedback during a task that would be completed on the second visit. Participants were shown up to five photographs of gender- and ethnicity-matched peers. They then selected three with whom they would like to interact, and rated the three peers for how likeable and attractive they looked on a scale from 1 (not very) to 10 (very). Participants also completed a personal survey where they listed information about themselves (birthday; hometown; and favorite music, TV shows, books, quoi5z9siandexianda1 37leryE(aqz9si3(aqz9si3(aqz9si3(nked)-16qz9snvpfy(looked)-169snvpf-160hy(First)-individuals)ber2-12(fst)216nquoi5z9



probability and time (pre-interaction, post-interaction) on likeability ratings ( $F_{(2,78)} = 5.48, < 0.01$ ; Fig. 2).  $P$  analyses indicated that post-task ratings decreased linearly with decreasing interaction probability, such that peers who interacted less with the participant were rated as less likeable (linear term  $F_{(1,39)} = 7.17, < 0.02$ ). Whereas pre-task likeability ratings were equivalent for all three peers ( $s > 0.48$ ), after the task the Frequent ( $F_{(39)} = -2.26, < 0.03$ ) and Continuous ( $F_{(39)} = -2.68, < 0.02$ ) peers were rated as more likeable than the rarely reinforcing peer, though there was not a significant difference in likeability ratings after the task between the Frequent and Continuous peers ( $F_{(39)} = -0.69, > 0.49$ ). Attractiveness ratings were not significantly modulated by task conditions (main effects of reinforcement probability, time, and interactions:  $s > 0.09$ ).

**A.** Participants responded correctly to 95.63% of trials (SD 3.54%). Response accuracy was not significantly modulated by the task conditions (main effects of reinforcement probability, time, and interactions:  $s > 0.29$ ).

Response latencies to the cue varied as participants learned the reinforcement contingency outcomes associated with each peer, as indicated by a significant interaction between probability of reinforcement and time ( $F_{(2,84)} = 3.98, < 0.03$ ; Fig. 2).  $P$  tests showed that, whereas there was no difference in reaction times in the early trials (all  $s > 0.41$ ), individuals were faster during the late trials for the Frequent reinforcement condition ( $F_{(42)} = 2.49, < 0.02$ ), compared with the Rare reinforcement condition. There was a trend for responses to be faster in the Continuous reinforcement condition ( $F_{(42)} = 2.01, < 0.06$ ) than in the Rare reinforcement condition. Overall, participants were faster during the late versus early trials ( $F_{(42)} = 15.21, < 0.01$ ) and there was no main effect of probability of social reinforcement when collapsing across time ( $F_{(2,84)} = 1.43, > 0.25$ ).

To further test for the effects of learning, we compared reaction times for Rare and Continuous reinforcement before and after reinforcement contingencies were reversed at the end of the experiment. Evidence that participants had implicitly learned the contingencies was further supported by the interaction between time (sixth run vs reversal) and reinforcement probability (rarely reinforcing vs continuously reinforcing) on reaction times ( $F_{(1,42)} = 10.15, < 0.01$ ; Fig. 2C).  $P$  tests showed a significant reaction time speeding when the Rare condition switched to delivering Continuous reinforcement ( $F_{(42)} = 3.13, < 0.01$ ). There were no main effects of reinforcement probability ( $F_{(5)} = 2.77, = 0.1$ ) or time ( $F_{(1,42)} = 1.35, > 0.25$ ). There was also no difference in the Frequent condition (unchanged during reversal) on reaction times between the last run and the reversal run ( $F_{(42)} = -1.48, > 0.15$ ).

Additionally, we examined how reaction times (RT) changed based on feedback from the preceding trial as another index of how the reinforcement contingencies altered behavior. We found that participants were faster on the subsequent trial after not receiving positive reinforcement (mean score RT:  $-0.01$ , SD: 0.13) versus when they had received positive reinforcement (mean -score RT: 0.07, SD: 0.14) ( $F_{(42)} = 2.86, < 0.01$ ).

## Imaging

$P$  As indexed by the prediction error parametric regressor, prediction error signals ( $\delta$ ) were positively associated with activity in the rostral anterior cingulate cortex, ventral striatum, anterior insula, and OFC (Table 1; Fig. 3). The parametric values in the general linear model encompassed positive and negative prediction errors, demonstrating that the BOLD fluctuations in these regions tracked learning signals reflecting reinforcement expectancies. Together, these findings delineate an orbital frontostriatal circuit showing significantly greater activity associated with the unexpected outcomes of either receiving or not receiving positive social reinforcement.

We also examined regions of the brain that positively correlated with learning to distinguish the faces of the peers based on their differential rates of positive social reinforcement (learned cue value). Specifically, group analysis of the cue phase of trials that tracked positively with modulations of expected value ( $\lambda$ ) identified greater activity in the rostral anterior cingulate cortex with larger expected value (Fig. 4). No other regions survived whole-brain correction. No regions within the frontostriatal circuitry of interest demonstrated negative correlations with expected value (i.e., brain regions sensitive to lower expected values) at corrected thresholding.

## Discussion

Repeated social exchanges shape our behavior toward others. In this experiment, we examined how different probabilities of positive interaction from distinct peers rapidly influence social learning. Within a reinforcement learning framework, we developed a novel social paradigm and demonstrated that the neural systems engaged while forming social expectations are similar to those involved in basic reward learning. This overlap in neural circuitry and function is consistent with prediction error-related learning and with our hypothesis that positive social interactions can serve as secondary reinforcers, taking on the attributes of primary re-

provide positive social reinforcement. This finding is similar to studies that demonstrate improved performance on a trial that follows receiving punishment (Hester et al., 2010) or choosing to make a bet more often after losing money than after winning money (Liu et al., 2007), though the present study did not assess strategic behavior directly. Together, the behavioral findings demonstrate that participants learned the reinforcement contingencies and thus provide an objective index of social learning.

The changes in likeability ratings and response latencies did not appear to be conscious behavioral choices. The majority (93%) of participants were unable to articulate the reinforcement patterns, suggesting little if any explicit awareness of the reinforcement contingencies. These findings demonstrate that social preferences and actions can be influenced after only brief encounters with peers and without conscious awareness. Such rapid changes highlight the influence of positive social interactions on effectively altering subsequent behavior.

The neural correlates of these behavioral changes draw upon the same neural circuitry as that implicated in reinforcement learning (Alexander et al., 1986; Haber and Knutson, 2010). Prediction error ( $\delta$ ) learning engaged the ventral striatum and orbital frontal cortex, similar to previous studies using single-cell recordings (Schultz et al., 1997; Fiorillo et al., 2003; Sul et al., 2010) and human imaging studies with primary reinforcers such as juice (McClure et al., 2003; O'Doherty et al., 2003; D'Ardenne

cety

ment outcomes. Specifically, one peer always provided positive social reinforcement, another one frequently provided positive social reinforcement, and the third rarely provided positive social reinforcement. Ratings of likeability changed from the beginning to the end of the experiment, with less reinforcing peers becoming less likeable, and more reinforcing peers yielding higher ratings of likeability by the end of the task. By asking participants to make a simple button response during the cue presentation, we tested whether speeding of response latencies (action tendencies) indexed learned associations between a given peer and their probability of providing positive social reinforcement. As expected due to the simplicity of the task, accuracy was at ceiling and there were no statistical differences in accuracy for the three peers.

In the current study, we observed faster responses to peers who provided positive social reinforcement more often, similar to studies where participants responded more quickly to cues that reliably predicted receiving a primary or secondary reward (O'Doherty et al., 2006; Spicer et al., 2007). Measuring differences in reaction times to cues to index learning differs from reinforcement studies that use modulated choice behavior as an indicator of learning (Tanaka et al., 2004; Daw and Doya, 2006; Schönberg et al., 2007). Choice tasks index changes in explicit preferences or a participant's strategy in maximizing reinforcement, while in the current study, changes in responses are thought to index differences in approach behaviors that are based on learning from a prior history of social feedback. In addition, participants showed faster reaction times after trials that did not

about investors' decisions (King-Casas et al., 2005; Phan et al., 2010). Our results complement these studies by demonstrating a neural mechanism for how prior positive interactions with others shape our expectations for future interactions. Given the increased sensitivity in the ventral striatum to appetitive stimuli during adolescence (Galvan et al., 2008; Somerville et al., 2011), as well as the greater influence of peers during adolescence (Spear, 2000; Gardner and Steinberg, 2005), this work clearly raises the question of how peer interaction differentially impacts learning and behavior across development and how this may be differentially represented in the brain. Accordingly, it would be interesting to explore whether adolescents show increased sensitivity during social learning relative to children and adults.

The expected values  $\downarrow$  to the cues corresponded with greater activity in the rostral anterior cingulate cortex. Previous studies have shown the rostral anterior cingulate cortex/medial prefrontal cortex is sensitive to cues that predict reward receipt (Tanaka et al., 2004; Knutson et al., 2005; Palminteri et al., 2009) and may play a role in general learning about the value of information and using this information for future decisions (Rushworth and Behrens, 2008). Lesion studies in nonhuman primates have shown this region is important for establishing patterns of social interest in other individual male or female macaques (Rudebeck et al., 2006). Human imaging studies have shown this region is sensitive when choosing to approach peers relative to celebrities (Güroglu et al., 2008) and when engaging in a series of actions during live relative to recorded interactions (Redcay et al., 2010). Given these studies examining social value in the anterior cingulate cortex, and the extensive literature showing a general sensitivity of this region in monitoring response conflict (Botvinick et al., 1999, 2004), our findings suggest that learning social cue values drives changes in behavior that may differ or conflict with the cognitive demands of the situation (e.g., task demands). Over the course of the experiment, this conflict may increase as behavior is modulated in response to changing expected values. Although the current study did not find that the orbital frontostriatal circuit was sensitive to expected values, the findings in the anterior cingulate cortex may suggest a role for this region in processing behavioral tendencies toward learned social cues. These findings thus offer insight into the neural processing of quick social decisions.

### Conclusions

Our findings provide direct evidence for how brief, positive, social interactions can significantly shape social learning across three discrete measures: social preferences, behavioral actions, and neural activity. After short interactions with others, social preferences and actions can be altered, highlighting the significance of social acceptance in biasing behavior. Moreover, we show that formal computational models of reinforcement learning apply to secondary reinforcement learning in the social domain. We demonstrate that the neural circuitry involved in forming prediction error signals about receiving social reinforcement, including the ventral striatum and orbital frontal cortex, overlaps with circuitry that subserves learning about other types of rewards (e.g., food or money). Overall, the findings suggest that similar mechanisms underlie basic reinforcement learning and our ability to rapidly and flexibly update our expectations during interactions with others, which enables us to effectively navigate the social environment.

### References

- Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9:357–381.
- Bandura A, Walters RH (1963) *Social learning and personality development*. New York: Holt, Rinehart and Winston.
- Baumeister RF, Leary MR (1995) The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull* 117:497–529.
- Berns GS, McClure SM, Pagnoni G, Montague PR (2001) Predictability modulates human brain response to reward. *J Neurosci* 21:2793–2798.
- Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402:179–181.
- Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn Sci* 8:539–546.
- Bray S, O'Doherty J (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol* 97:3036–3045.
- Cox RW (1996) AFNI: Software for analysis and visualization of functional

- Li J, Daw ND (2011) Signals in human striatum are appropriate for policy update rather than value prediction. *J Neurosci* 31:5504–5511.
- Li J, Delgado MR, Phelps EA (2011) How instructed knowledge modulates the neural systems of reward learning. *Proc Natl Acad Sci U S A* 108:55–60.
- Lin A, Adolphs R, Rangel A (2011) Social and monetary reward learning engage overlapping neural substrates. *Soc Cogn Affect Neurosci*. Advance online publication. Retrieved March 22, 2011. doi:10.1093/scan/nsr006.
- Liu X, Powell DK, Wang H, Gold BT, Corbly CR, Joseph JE (2007) Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. *J Neurosci* 27:4587–4597.
- McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in