



Vicarious reinforcement in rhesus macaques (*Macaca mulatta*)

Steve W. C. Chang^{1,2*}, Amy A. Winecoff¹ and Michael L. Platt^{1,2,3,4}

¹ Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

² Department of Neurobiology, Duke University, Durham, NC, USA

³ Department of Evolutionary Anthropology, Duke University, Durham, NC, USA

⁴ Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

Edited by:

Daeyeol Lee, Yale University School of Medicine, USA

Reviewed by:

Mehrdad Jazayeri, University of Washington, USA

James Rilling, Emory University, USA

*Correspondence:

Steve W. C. Chang, Center for Cognitive Neuroscience, B203 Levine Science Research Center, Duke University, Box 90999, Durham, NC 27708, USA.

e-mail: steve.chang@duke.edu

What happens to others profoundly influences our own behavior. Such other-regarding outcomes can drive observational learning, as well as motivate cooperation, charity, empathy, and even spite. Vicarious reinforcement may serve as one of the critical mechanisms mediating the influence of other-regarding outcomes on behavior and decision-making in groups. Here we show that rhesus macaques spontaneously derive vicarious reinforcement from observing rewards given to another monkey, and that this reinforcement can motivate them to subsequently deliver or withhold rewards from the other animal. We exploited Pavlovian and instrumental conditioning to associate rewards to self (M1) and/or rewards to another monkey (M2) with visual cues. M1s made more errors in the instrumental trials when cues predicted reward to M2 compared to when cues predicted reward to M1, but made even more errors when cues predicted reward to no one. In subsequent preference tests between pairs of conditioned cues, M1s preferred cues paired with reward to M2 over cues paired with reward to no one. By contrast, M1s preferred cues paired with reward to self over cues paired with reward to both monkeys simultaneously. Rates of attention to M2 strongly predicted the strength and valence of vicarious reinforcement. These patterns of behavior, which were absent in non-social control trials, are consistent with vicarious reinforcement based upon sensitivity to observed, or counterfactual, outcomes with respect to another individual. Vicarious reward may play a critical role in shaping cooperation and competition, as well as motivating observational learning and group coordination in rhesus macaques, much as it does in humans. We propose that vicarious reinforcement signals mediate these behaviors via homologous neural circuits involved in reinforcement learning and decision-making.

Keywords: vicarious reinforcement, social reward, gaze, social interaction, rhesus macaques

INTRODUCTION

Reinforcement learning provides a powerful mechanism for associating stimuli and actions with the direct experience of reward and punishment (Rescorla and Wagner, 1972; Schultz et al., 1997; Sutton and Barto, 1998). Behavioral and neurobiological evidence indicate that human behavior also depends on outcomes that have not been directly experienced. For example, fictive, or counterfactual, learning describes sensitivity to reward outcomes for options that were not chosen, were merely observed, or were even imagined (Byrne, 2002; Lohrenz et al., 2007; Epstude and Roese, 2008). Fictive learning can be described formally in terms analogous to reinforcement learning (Lohrenz et al., 2007), and may depend on overlapping neural circuitry (Lohrenz et al., 2007; Hayden et al., 2009; Mobbs et al., 2009).

Observing what happens to others also powerfully shapes human learning and behavior (Berger, 1962; Bandura and McDonald, 1963; Bandura et al., 1963). Such other-regarding outcomes can drive observational learning (Mobbs et al., 2009; Jeon et al., 2010), and motivate other-regarding behaviors such as cooperation and charity, as well as spite and schadenfreude (Takahashi et al., 2009). The “warm glow” hypothesis (Andreoni, 1990) suggests that vicarious reward and punishment motivates individuals to prefer either

positive or negative outcomes to others (Bandura et al., 1963; Fehr and Fischbacher, 2003; Mobbs et al., 2009). Human social emotions associated with vicarious reward and punishment, such as fairness and envy, appear early in ontogeny, and their derangement in mental disorders like psychopathy can have devastating consequences (Kiehl, 2006).

Such observations endorse the idea that neural mechanisms supporting vicarious reinforcement are derived specializations of the human brain, which support complex social behavior including observational learning, cooperation, and even altruism (Fehr and Fischbacher, 2003). Though highly specialized for complex social behavior in humans, these mechanisms appear to have deep evolutionary roots. Behavioral and neurobiological evidence demonstrate rudimentary forms of fictive, observational, and social learning in non-human animals. Rhesus macaques, for example, learn from fictive outcomes and this process appears to be supported by the same circuitry mediating fictive learning in humans (Hayden et al., 2009). In some species, learning to perform a task is facilitated by watching others learn the same task (Zentall and Levine, 1972; Zentall et al., 1996; Drea and Wallen, 1999; Subiaul et al., 2004; Whiten et al., 2009). Chimpanzees are capable of learning to use complex tools by observing others (Tomasello et al.,

1987), and their observational learning seems to be contingent on the associative strength of observed action and outcome (Crawford and Spence, 1921). Observing another mouse receive a shock can drive fear conditioning in the observer, and this observational fear conditioning depends on affective pain circuitry that has been implicated in empathy in humans (Jeon et al., 2010).

Whether mere observation of rewarding events occurring to another individual can drive the expression of social preferences in non-human animals, as proposed by the “warm glow” model, however, remains debated. Some have argued that the expression of other-regarding preferences in humans reflects the evolution of mechanisms that promote cooperative reproduction, but the evidence for other-regarding behaviors in cooperatively breeding animals remains controversial (Burkart et al., 2007; de Waal et al., 2008; Lakshminarayanan and Santos, 2008; Massen et al., 2010). Others have argued that only those species most closely related to humans, namely chimpanzees and bonobos, possess the derived features of human biology and cognition, in particular “theory of mind” (Call and Tomasello, 2008), express other-regarding preferences, but again the evidence for such behavior in apes remains inconclusive (Tomasello et al., 2003; Silk et al., 2005).

We hypothesize instead that cooperation and competition endemic to group life favors the evolution of neural circuits tuned to extract information about the experiences of others, and that these circuits serve as the core building blocks for the development of observational learning and other-regarding behaviors, which reach their fullest expression in our own species. As a first behavioral test of this idea, we probed the impact of vicarious reinforcement on subsequent decisions made by rhesus macaques with respect to other monkeys. Rhesus monkeys observe others to gather social information (Cheney and Seyfarth, 1990), display sensitivity to fictive outcomes in non-social settings (Hayden et al., 2009), show rudimentary understanding of the intentions of others (Flombaum and Santos, 2005), care for kin (Maestripieri, 1994), and may give up foods to alleviate pain in conspecifics (Masserman et al., 1964). We hypothesized that such behaviors, as well as naturally occurring behaviors such as social grooming, alliance formation, and group territorial defense, derive from fundamental vicarious reinforcement mechanisms similar to those guiding social behavior in humans.

To test this hypothesis, we capitalized on simple Pavlovian and instrumental conditioning to associate liquid rewards to self and rewards to another monkey with a set of visual cues, and subsequently tested for preferences amongst these cues in a two alternative-forced choice task to infer underlying reward associations. Subsequent preference tests between cues revealed a preference to reward the other monkey rather than no one, but a preference to withhold reward from the other when choosing between rewarding self or both monkeys simultaneously. Crucially, monkeys showed no preferences amongst the cues when the other monkey was removed from the room and replaced with a juice collection bottle, confirming the social dependence of vicarious reinforcement and thus ruling out simple fictive learning as an explanation for the observed behavior. Preferences amongst cues were predicted by the relative subjective value of each cue, as inferred from the time it took to initiate choosing each option, as well as the frequency with which the actor monkey looked at the recipient monkey following

choices. These findings demonstrate context-dependent, vicarious reinforcement guides decision-making with respect to others in rhesus macaques.

MATERIALS AND METHODS

GENERAL PROCEDURES

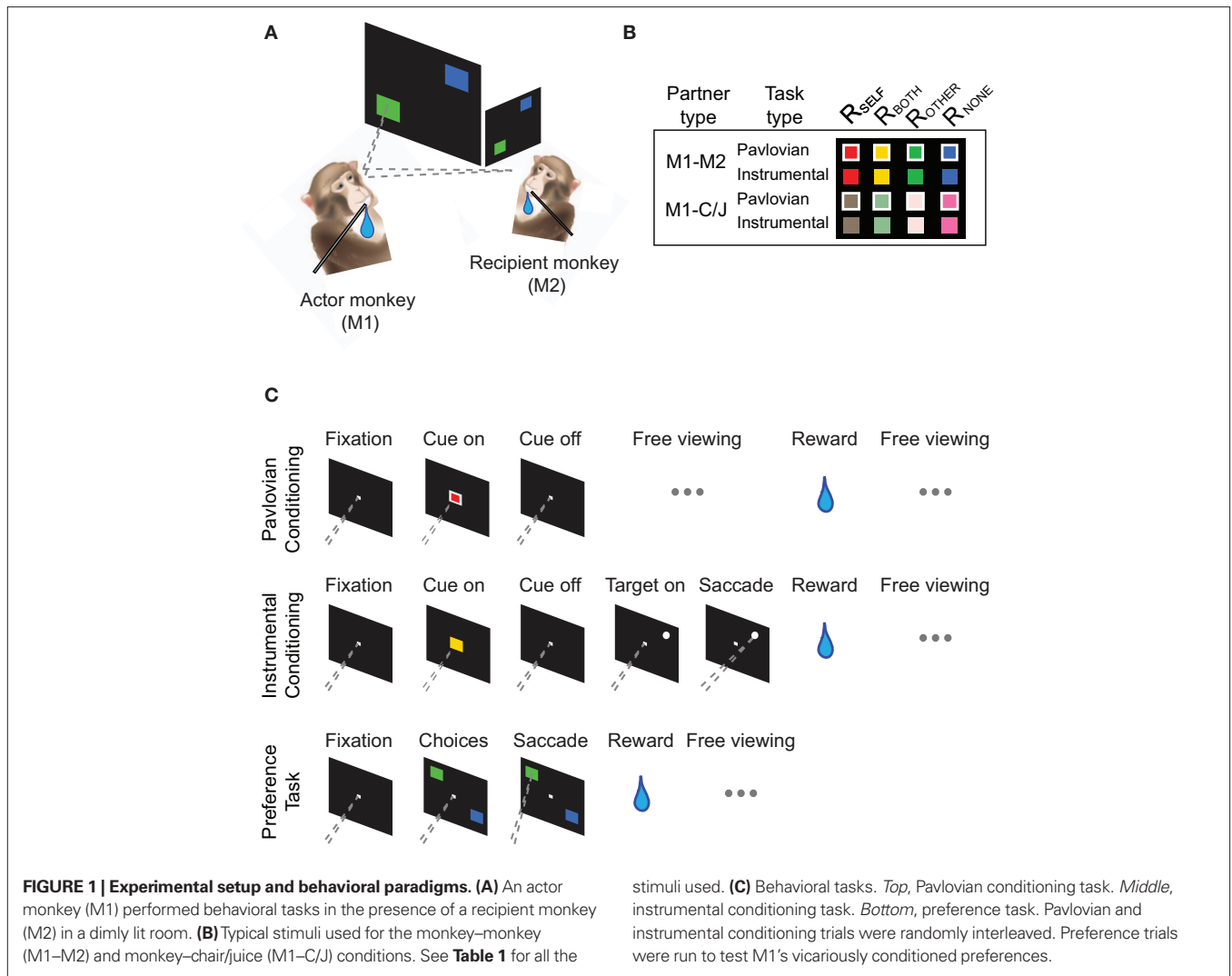
All procedures were approved by the Duke University Institutional Animal Care and Use Committee and were designed and conducted in compliance with the Public Health Service’s Guide for the Care and Use of Animals. All rhesus macaques (*Macaca mulatta*) used in the study were genetically unrelated, middle-ranked males (mean age and SD, 9 ± 3.7), and none of M1–M2 pairs were cagemates. All monkeys involved in this study received at least 20 ml/kg of liquid daily in addition to fluid earned in the experiment.

Horizontal and vertical eye positions were sampled (1000 Hz) by an infrared eye-monitoring camera system (SR Research Eyelink). Stimuli were controlled by a computer running Matlab using PsychToolbox (Brainard, 1997; Pelli, 1997). All experiments were carried out in a dimly lit room to ensure visibility of M1 and M2. Both M1 and M2 were head-restrained during the experiments. M2 was always situated diagonally across from M1 at a 45° eccentricity to the right from the center of M1’s screen, and they faced each of their own display screens, which were located at a 90° angle from one another (Figure 1A). The location of M2 (center of the face) was mapped empirically prior to experiments using M1’s eye positions. In the chair/juice control, an empty primate chair with an operating juice tube and a depository bottle replaced M2. The depository bottle was placed in the same space that would otherwise be occupied by M2’s mouth region (all else in the control were identical to the M1–M2 condition).

Solenoid valves that delivered the liquid rewards were placed in another room to prevent monkeys from forming secondary associations between solenoid clicks and different reward types. We also included a separate solenoid designated for R_{NONE} that only produced clicks but delivered no fluid. Masking white noise was always played in the experimental room. We used a relatively large juice reward size (0.5–1 ml) per successful trial in order to clearly demonstrate to M1 that M2 received juice rewards on R_{BOTH} and R_{OTHER} trials. The reward size remained constant across different reward conditions within each block. More specifically, the fluid-restricted actor and recipient monkeys received, on average, 250 ml of liquid in the form of cherry juice. The amount of fluid intake across different experimental sessions only fluctuated within ~50 ml. During the days without experimental sessions, the monkeys drank up to 500 ml *ad lib.* or more, which demonstrates the high motivational level. Furthermore, they were very motivated by this reinforcement schedule, given that they participated in the experiments and continued to perform trials for about 2–3 h without stopping.

BEHAVIORAL TASKS AND ANALYSIS

The behaviors from two actor monkeys were examined. The tasks were initially developed for neurophysiological investigations, and therefore we limited the number of the actor monkeys to 2, which is the standard and practical convention for neurophysiological studies. This convention, however, weakens the generalizability of the study. To address this to our best, the current study also reports the main findings and statistics separately for the two monkeys. A



total of eight M1–M2 and two M1–chair/juice pairs were used in the study. Of these, three M1–M2 pairs and two chair/juice controls (for each M1) were subjected to both Pavlovian and instrumental conditioning trials with a novel stimulus set for each pair. The remaining five M1–M2 pairs were tested based on already learned cue–reward associations from the conditioning trials (i.e., from the three M1–M2 pairs). The complete set of visual cues used is shown in **Table 1**. One actor monkey (MY) served as M1 first then was also tested as M2 at the very end, whereas the other actor monkey (MO) was tested as M2 at the very beginning, then served as M1 from then on (see box in **Figure 3B** for the complete list of pairings). Context-dependent preferences were evident in both M1s (see main text for statistics). Other monkeys involved in the study only served as M2.

The conditioning task consisted of randomly interleaved Pavlovian (**Figure 1B**, top) and instrumental conditioning trials (**Figure 1B**, middle). On both trial types, M1 initiated the trial by shifting gaze to a central stimulus ($0.7^\circ \times 0.7^\circ$). After 200 ms of fixation, a cue ($5^\circ \times 5^\circ$) of different shape and/or color appeared in the center and remained on for 1 s on Pavlovian trials and for 300 ms on instrumental trials. Visual cues on Pavlovian trials

contained a white outline around the same cues used to convey the same reward outcomes on instrumental trials (**Figure 1C**). On Pavlovian trials, cue onset marked the end of the fixation requirement (i.e., free to look anywhere), and the appropriate reward outcome was delivered. On instrumental trials, however, extinction of the cue was followed by another 200 ms of central fixation before a white target stimulus (1° diameter) appeared at one of eight random locations (eccentricity of 8°). M1 had 1.5 s to shift gaze to the target with in 3.4° . After successful target acquisition, the appropriate reward was delivered. At the onset of reward, M1 was free to look anywhere in the setup before the next trial began for 1 s. Rewards were delivered at approximately the same time for the Pavlovian and instrumental trials after matching the reward timings of the previously occurred instrumental trials (requiring motor responses) to the subsequent Pavlovian trials on a trial-by-trial basis. Data from 120 ± 57 (median \pm SD) and 173 ± 59 correct trials were collected for each pair and the non-social control, respectively.

In the preference task (**Figure 1C**, bottom), M1 again began each trial by shifting gaze to the fixation stimulus. After 200 ms of central fixation, two of the previously learned cues from the

Table 1 | Stimulus–reward pairs used in the experiments.

Stimulus–reward associations				Conditioned pairs on the conditioning trials (M1–M2)	Preference tested pairs on the preferential trials (M1–M2)		
R_{SELF}	R_{BOTH}	R_{OTHER}	R_{NONE}				
					MY–MD	MO–MD	MO–MB
				MO–MD	MY–ML	MO–ML	
				MO–MY	MY–MO	MO–MS	
				MO–MY		MO–MY	
				MY–C/J		MY–C/J	
				MO–C/J		MO–C/J	

Stimuli used for all individual monkey–monkey (e.g., MY–MD) pairs and monkey–chair/juice (e.g., MY–C/J) controls. On Pavlovian trials, a white outline was present on these cues (e.g., see **Figure 1C**).

conditioning task appeared as targets at two of eight random locations 8° from the central fixation stimulus, separated by 180° (e.g., **Figures 1A,B**, bottom). Upon target onset, M1 shifted gaze to one or the other target, and the reward outcome associated with that chosen target was delivered. M1 had 1.5 s to shift gaze to the target ($\pm 3.4^\circ$). Data from 229 ± 88 and 122 ± 78 correct trials were collected for each pair and the non-social control, respectively.

For both tasks, when an error occurred (i.e., failure to maintain fixation after cue onset or inaccurate gaze shift to the peripheral target), the trial was aborted, and a white error square ($14.2^\circ \times 14.2^\circ$) appeared on the screen for 1.5 s. On Pavlovian conditioning trials, errors were defined as failures to maintain fixation after acquiring the fixation point to start a trial. Because these errors were independent of any reward contingencies (i.e., before cue onset), we did not consider them here. On instrumental conditioning and choice trials, errors were defined as either failures to maintain fixation in the beginning of a trial or breaking fixation or not acquiring a target after the reward contingencies were revealed (after cue onset). In practice, almost all errors resulted from monkeys looking up and away from the computer monitor. Error trials were excluded from further analyses.

We calculated a vicarious reinforcement index (VRI) by computing the difference between the frequency of choosing one option (n_A) and the other (n_B) and then normalizing the difference by the sum:

$$\text{VRI} = \frac{n_A - n_B}{n_A + n_B} \quad (1)$$

In the Self/Both context, n_A and n_B were the number of R_{BOTH} and R_{SELF} choices, respectively, whereas in the Other/None context, they were R_{OTHER} and R_{NONE} , respectively. The VRI always ranged from -1 to 1 , with 1 corresponding to M1 always choosing the prosocial option (either R_{BOTH} or R_{OTHER}), -1 corresponding to M1 always choosing the non-prosocial option (either R_{SELF} or R_{NONE}), and 0 corresponding to M1 choosing each of the alternatives equally often.

Saccade reaction times (RTs; time from target onset to movement onset) were computed using a $20^\circ/\text{s}$ velocity crossing threshold on each trial. The frequency of M1 looking at M2 was computed by counting the number of gaze shifts made by M1 into a $25^\circ \times 25^\circ$

window spanning from the center of M2's face during the pre-reward free-viewing period (from the start of reward delivery up to 1 s after reward completion of the delivery; **Figure 1B**). On non-social control trials, this region was occupied by an operating juice tube and a depository bottle situated in the neckplate of our primate chairs.

RESULTS

MONKEYS EXHIBIT VICARIOUS REINFORCEMENT

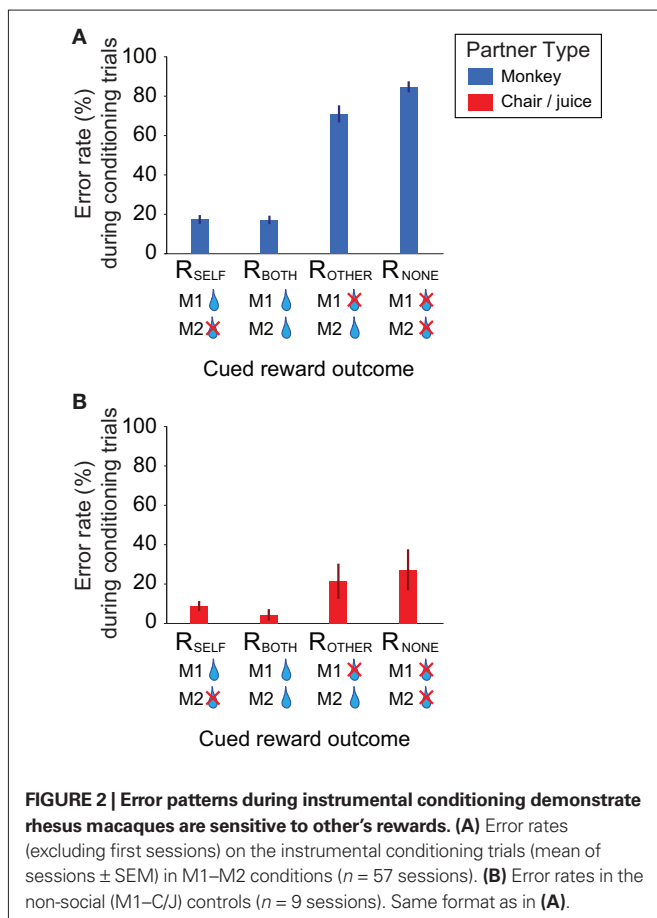
Two adult male rhesus monkeys served as actors (M1) and five adult male rhesus monkeys served as recipients (M2; see Materials and Methods). M1 and M2 sat across from each other (**Figure 1A**), and each viewed his own computer screen, which displayed visual cues. On Pavlovian trials (**Figure 1B**, top), M1 and M2 both saw the same cue at the center of the display, and juice rewards were delivered to M1 (R_{SELF}), M2 (R_{OTHER}), both M1 and M2 (R_{BOTH}), or neither (R_{NONE}) depending on the color or shape of the cue (**Figure 1C**; **Table 1**). On instrumental trials (**Figure 1B**, middle), M1 and M2 again both saw the same cue and a neutral target appeared, to which M1 had to shift gaze for subsequent delivery of juice reward to M1, M2, both M1 and M2, or neither.

Error rates (failure to maintain fixation after cue onset or inaccurate gaze shift to the peripheral target; see Materials and Methods) on the instrumental trials demonstrated that M1 discriminated among the four reward conditions (**Figure 2A**). For both instrumental conditions in which M1 received direct fluid reward, error rates were indistinguishable whether or not M2 was also rewarded (R_{SELF} and R_{BOTH} ; $p = 0.54$, Wilcoxon sign rank test; $n = 57$ sessions; **Figure 2A**). In contrast, M1 made significantly more errors on trials with cues that did not result in direct fluid reward to self compared with trials displaying cues that predicted direct fluid reward to self (R_{SELF} or R_{BOTH} versus R_{OTHER} or R_{NONE} ; all $p < 0.00001$, Wilcoxon sign rank test; $n = 57$ sessions; **Figure 2A**).

Notably, M1 continued to perform instrumental trials with cues predicting reward to M2 (R_{OTHER}) or no one (R_{NONE}) despite the fact that he was never rewarded in either case and error rates clearly showed that M1 did not prefer these cues [error rates: $71.0 \pm 4.3\%$ (mean \pm SEM per session) and $84.8 \pm 2.8\%$, respectively]. Critically, M1 made significantly fewer errors when the cue predicted a fluid reward for M2 compared with when

the cue predicted no one would receive a fluid reward, indicating a reinforcing property to observing M2 receive a reward ($p < 0.00001$, Wilcoxon sign rank test; $n = 57$; **Figure 2A**). This pattern of systematically lower error rates on R_{OTHER} compared to R_{NONE} was also evident in each M1 individually (both $p < 0.005$, Wilcoxon sign rank test; $n = 44$ for MY and 13 sessions for MO). In contrast, in a non-social control in which M2 was replaced with a collecting bottle (chair/juice control; see Materials and Methods), the error rates for responding to cues predicting reward to other (R_{OTHER}) and reward to no one (R_{NONE}) were statistically indistinguishable ($p = 0.20$, Wilcoxon sign rank test; $n = 9$ sessions; **Figure 2B**).

The presence of another monkey clearly influenced error rates during conditioning. M1 made fewer errors overall in the non-social control compared to the social trials (total error rates: $15.5 \pm 3.8\%$ versus $47.6 \pm 2.5\%$, $p < 0.00001$, Wilcoxon rank sum test; $n = 57$; **Figure 2A**). The higher error rates on the social compared to the non-social control trials could be attributed to increased attentional demands due to the presence of another monkey (e.g., bystander effect). Error rates during the conditioning trials demonstrate that rhesus monkeys value rewards to self more than they value rewards to others, as expected. Nonetheless, the fact that M1 continued to participate when only M2 was rewarded directly with juice suggests that observing another monkey receive a reward is vicariously reinforcing.



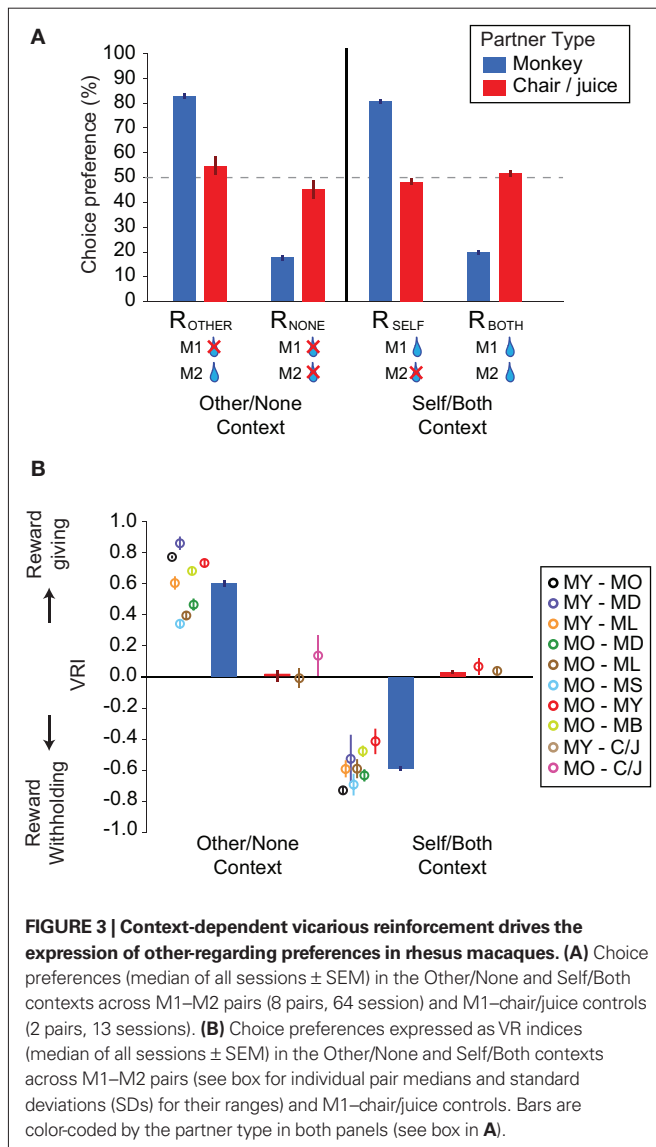
CONTEXT-DEPENDENT MANIFESTATION OF VICARIOUS REINFORCEMENT

Subsequently, we used a two alternative forced task (preference task; **Figure 1B** bottom) to directly test the hypothesis that observing another monkey receiving a reward is vicariously reinforcing. In the preference task, M1 chose between pairs of previously conditioned cues (R_{SELF} versus R_{BOTH} or R_{OTHER} versus R_{NONE}) by shifting gaze to one of them. Critically, rewards were matched between the available choices in each condition – that is, M1 chose between R_{OTHER} and R_{NONE} [Other/None condition (purely vicarious context); M1 never directly rewarded with juice] or between R_{BOTH} and R_{SELF} (Self/Both condition; M1 always rewarded with juice). We hypothesized that cues would acquire value vicariously via Pavlovian and instrumental conditioning, and that differential cue values would be expressed as systematic preferences in this choice task.

As expected, error rates in the preference task were consistent with a preference for receiving direct fluid reward in the Self/Both condition (error rate: $0.8 \pm 0.2\%$; $n = 64$ sessions), compared to no fluid reward in the Other/None condition, in which M1 was never rewarded ($12.6 \pm 1.8\%$; $p < 0.00001$, Wilcoxon sign rank test; $n = 64$). Remarkably, however, M1 performed about 88% of trials in which he was not directly rewarded with fluid. Again, as in the conditioning trials, M1 made significantly fewer errors during the non-social control ($n = 13$ sessions) compared to when M2 was present ($p < 0.001$, Wilcoxon rank sum test). M1 was significantly more willing to complete trials which resulted in no reward to M1 during the preference trials compared to the Pavlovian conditioning trials (correct rate: $87.4 \pm 1.8\%$ versus $22.1 \pm 2.7\%$, $p < 0.00001$, Wilcoxon rank sum test). This is consistent with prior observations in rhesus macaques that voluntary choices are more motivating than simple operant responses in the conditioning tasks (Suzuki, 1999).

The critical question was whether M1 acquired an intrinsically rewarding preference, through vicarious reinforcement, for rewarding M2 in the absence of rewarding self (R_{OTHER}). The choice preferences of M1 demonstrated that cues indeed acquired strong motivational associations even when M1 received no direct reward. M1 consistently preferred R_{OTHER} ($82.5 \pm 1.1\%$) over R_{NONE} (17.5% ; $p < 0.00001$, Wilcoxon signed rank test; $n = 64$ sessions), even though M1 was never directly rewarded with juice in this context (**Figure 3A**). Critically, this preference was absent in the non-social control when M2 was removed from the experimental room and replaced by an operating juice tube and a collection bottle [**Figure 3A**; 54.7 ± 3.8 (R_{OTHER}) versus 45.3% (R_{NONE}), $p = 0.17$, Wilcoxon sign rank test; $n = 13$]. In contrast, in the Self/Both context, M1 consistently preferred R_{SELF} ($80.3 \pm 1.0\%$) over R_{BOTH} (19.7% ; $p < 0.00001$, Wilcoxon sign rank test; $n = 64$), even though either choice led to the same physical juice reward for M1 simultaneously (**Figure 3A**). This pattern was again absent in the non-social control [**Figure 3A**; 48.3 ± 1.3 (R_{SELF}) versus 51.7% (R_{BOTH}), $p = 0.06$, Wilcoxon sign rank test; $n = 13$]. We observed the context-dependent patterns of behavior in each M1 separately [percentage of choosing R_{OTHER} (MY and MO): 86.7 ± 1.5 and $80.7 \pm 1.4\%$; percentage of choosing R_{SELF} : 85.4 ± 1.6 and $79.6 \pm 1.1\%$].

We further quantified M1's preferences by calculating a VRI, a contrast ratio varying from -1 to 1 , with positive values indicating preferences for R_{OTHER} over R_{NONE} (Other/None condition)



or R_{BOTH} over R_{SELF} (Self/Both condition) and 0 indicating indifference (see Materials and Methods). Analysis of the index led to similar results. In the Other/None condition, M1 preferred to reward M2 (VRI: 0.60 ± 0.02 , significantly different from 0, $p < 0.00001$, Wilcoxon sign rank test; $n = 64$ sessions; **Figure 3B**), and this pattern was absent in the non-social control (0.11 ± 0.08 , $p = 0.18$, Wilcoxon sign rank test; $n = 13$; **Figure 3B**). In the Self/Both context, however, M1 preferred to withhold reward from M2 (-0.58 ± 0.02 , $p < 0.00001$, Wilcoxon sign rank test; **Figure 3B**), and this pattern was only weakly evident in the non-social control (0.06 ± 0.03 , $p = 0.06$, Wilcoxon sign rank test; **Figure 3B**). Again, we observed the same pattern in each M1 separately [Other/None context (MY and MO): 0.70 ± 0.03 and 0.56 ± 0.03 ; Self/Both context: -0.65 ± 0.03 and -0.55 ± 0.02 ; all $p < 0.00001$, Wilcoxon sign rank test; $n = 19$ and 45, respectively]. These preferences remained stable over the course of data collection (**Figure 4**). Crucially, the VRI indices in the Other/None and Self/Both contexts never crossed over.

SOCIAL VARIABLES INFLUENCE VICARIOUS REINFORCEMENT

The magnitudes of the VRI were idiosyncratic to individual pairs of monkeys. Such differences were apparent from the very beginning of testing and remained more or less stable (**Figure 4**). We tested whether a specific social variable could explain this individual variability. First, we examined social status, which is known to influence social behaviors in both young children and non-human animals (Hawley, 1999), and observational learning has been implicated in how monkeys acquire social hierarchical information (Cheney and Seyfarth, 1990). We found that M1 was more willing to share reward if M1 was dominant to M2 in the Self/Both context ($n = 4$ out of 8). Specifically, M1 was more likely to choose R_{BOTH} in the Self/Both context [VRI: -0.54 ± 0.03 (M1 is dominant) versus -0.65 ± 0.03 (M1 is subordinate), $p < 0.01$, Wilcoxon rank sum test], but not necessarily R_{OTHER} in the Other/None context (0.62 ± 0.02 versus 0.58 ± 0.04 , $p = 0.57$, Wilcoxon rank sum test), if M1 is dominant to M2.

Second, we examined whether the familiarity of individuals in each pair biased choices by analyzing the housing locations of M1 relative to M2 in the colony room, which served as our measure of familiarity. It has been documented that social interaction behaviors increase with familiarity in both humans and monkeys (Preston and de Waal, 2002). We therefore reasoned that monkeys who could directly view each other (housed on opposite sides, compared to on same sides) would be more familiar and thus more likely to reward others. We found that VRI in the Other/None context was higher if M1 and M2 were housed on opposite sides ($n = 4$ out of 7) of the colony room, with direct visual access to each other. That is, M1 was more likely to choose R_{OTHER} in the Other/None context [0.71 ± 0.02 (opposite side) versus 0.53 ± 0.03 (same side), $p < 0.0001$, Wilcoxon rank sum test], but not necessarily R_{BOTH} in the Self/Both context (-0.60 ± 0.03 versus -0.57 ± 0.02 , $p = 0.19$, Wilcoxon rank sum test), if he could see him while in his home cage. Together, these findings suggest that individual variability in vicarious reinforcement (**Figures 3B and 4**) is at least partially influenced by both social dominance and social familiarity, although our limited sample size and types preclude strong conclusions.

MONKEYS OBSERVE THE REWARDING EVENTS OF OTHERS

After monkeys expressed their choice, they were permitted to freely look about (**Figure 1B**). During this free-viewing period, M1 often shifted gaze toward the face of M2, and the overall rate of shifting gaze depended on the reward outcome for M1 [**Figure 5A**; $20.5 \pm 3.6\%$ (median \pm SEM of the average between R_{OTHER} and R_{NONE}) versus $3.0 \pm 2.5\%$ (R_{SELF} and R_{BOTH}), $p < 0.00001$, Wilcoxon sign rank test; $n = 64$ sessions]. Critically, however, M1 looked at M2 more frequently after choosing to reward him over no one in the Other/None condition (R_{OTHER} : $25.4 \pm 4.1\%$; R_{NONE} : $16.7 \pm 5.8\%$, $p < 0.0005$, Wilcoxon sign rank test). We found a significant effect (frequency of gaze after choosing $R_{\text{OTHER}} > R_{\text{NONE}}$) in both M1s separately [MY: 23.1 ± 1.6 versus $15.4 \pm 2.3\%$ ($p < 0.005$; $n = 19$); MO: 26.1 ± 5.6 versus $17.9 \pm 8.1\%$ ($p < 0.01$; $n = 45$), Wilcoxon sign rank test]. Thus, our observation confirms that there is a link between social attention and vicarious reinforcement.

By contrast, in the non-social condition ($n = 13$ sessions), looking behavior was greatly reduced across all reward outcomes, compared to the social conditions ($R_{\text{SELF}} R_{\text{OTHER}}$: $p < 0.01$; R_{BOTH} : $p = 0.12$; R_{NONE} :

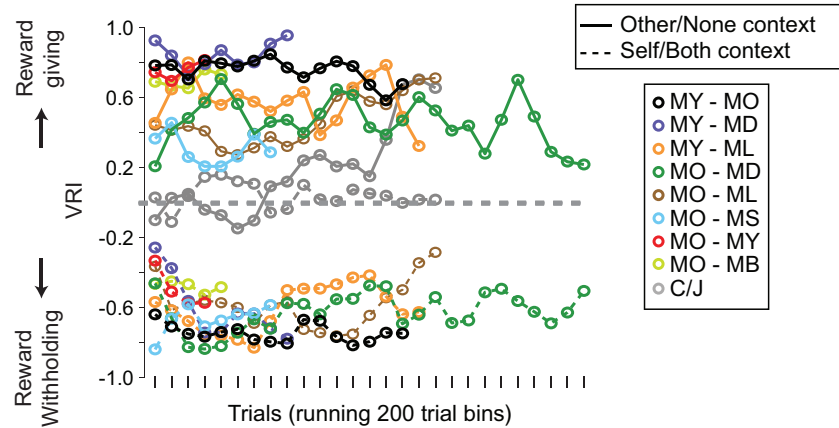


FIGURE 4 | Temporal progression (moving 200-trial bins with a step size of 100 trials) of context-dependent VR indices for individual M1–M2 pairs (8 pairs, 64 sessions) and M1–C/J pairs (2 pairs, 13 sessions; see box for pair identities). Data points on the right show individual pair medians and SDs across all trial bins for each pair.

$p = 0.01$, Wilcoxon rank sum test; **Figure 5A**). Critically, M1 neither looked at the juice bottle more often after choosing R_{OTHER} over R_{NONE} in the Other/None condition ($p = 0.34$, Wilcoxon sign rank test), nor after choosing R_{BOTH} over R_{SELF} in the Self/Both condition ($p = 0.94$, Wilcoxon sign rank test). The only factor that explained looking behavior in the non-social control was whether or not M1 was directly rewarded with juice (R_{SELF} and R_{BOTH} versus R_{OTHER} and R_{NONE} , $p < 0.0005$, Wilcoxon rank sum test). Thus, reward consumption by another monkey strongly recruits attention in the absence of direct reward to self, suggesting vicarious reinforcement may be mediated by social attention circuits in the brain (Klein et al., 2009).

SACCADE REACTION TIMES REVEAL THE INTERNAL DELIBERATIVE PROCESS

The pattern of saccade RTs on choice trials further corroborates the hypothesis that rewarding self was more valued than any other alternatives (**Figure 5B**; RTs for $R_{\text{SELF}} < R_{\text{BOTH}} < R_{\text{OTHER}} < R_{\text{NONE}}$; all comparisons $p < 0.00001$, Wilcoxon sign rank; $n = 64$ sessions). Generally, M1 responded more quickly whenever he chose to directly reward himself with juice. Nonetheless, M1 responded faster when he chose to reward M2 than when he chose to reward no one at all. These results were obtained for each M1 separately (all comparisons $p < 0.01$ for each M1, Wilcoxon sign rank test; $n = 19$ and 45 sessions for MY and MO, respectively). Importantly, in the absence of M2 (non-social control; $n = 13$ sessions), RTs across different reward outcomes remained more or less flat (**Figure 5A**). RTs were indeed slower overall in the presence of M2, perhaps due to an additional attentional load induced by the presence of M2 (blue versus red traces in **Figure 5B**; all comparisons $p < 0.005$, except $p = 0.46$ for R_{SELF} conditions, Wilcoxon rank sum test).

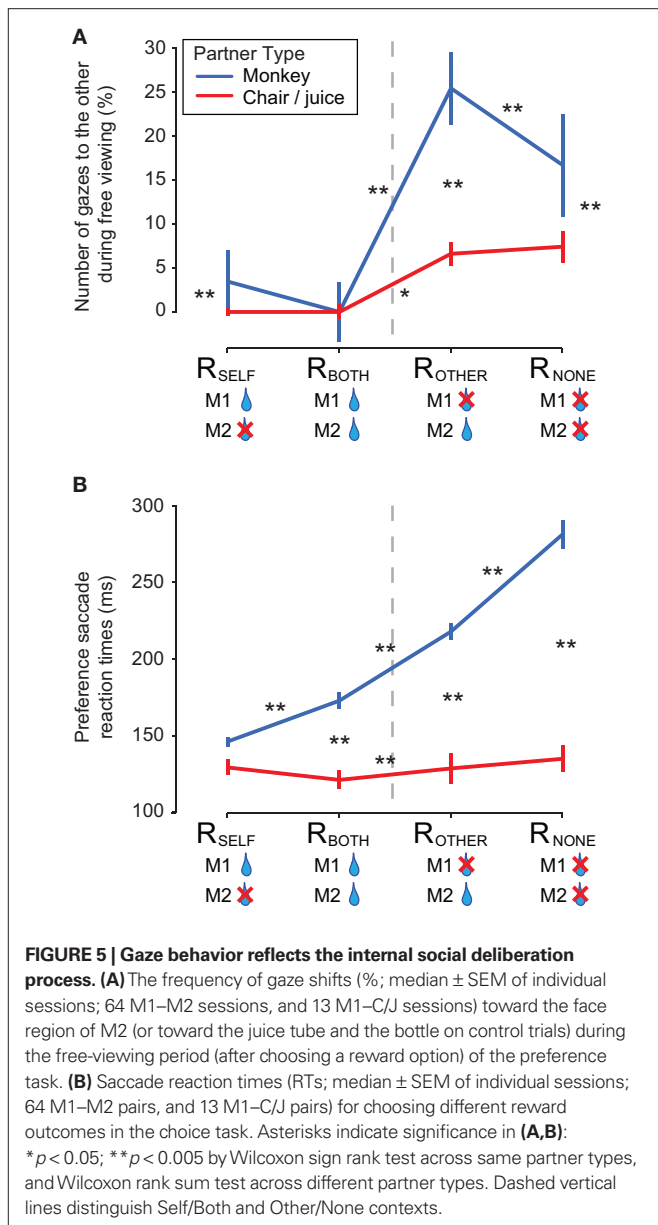
Given that monkeys generally respond more slowly when they anticipate smaller rewards (Kawagoe et al., 1998; Roesch and Olson, 2004), we inferred the subjective reward value of the four conditions to be $R_{\text{SELF}} > R_{\text{BOTH}}$ and $R_{\text{OTHER}} > R_{\text{NONE}}$. These inferred subjective reward values, which were absent in the non-social control (**Figure 5B**), predict the relative preferences between cues observed in the preference task (**Figure 3**). Specifically, M1 chose R_{SELF} over

R_{BOTH} in the Self/Both condition and showed faster RT for choosing R_{SELF} whereas M1 chose R_{OTHER} over R_{NONE} in the Other/None condition and showed faster RT for choosing R_{OTHER} .

DISCUSSION

We demonstrated that social preferences of rhesus macaques – non-human primates that live in large, hierarchical, mixed-sex social groups and who last shared a common ancestor with humans some 25 million years ago – could be shaped by vicarious reinforcement in a context-specific manner. Monkeys systematically preferred to provide juice reward to others rather than to no one, as if observing others drink is vicariously rewarding. In contrast, monkeys systematically withheld reward from others when confronted with the options to either consume reward alone or share reward. Increased social attention to M2 (i.e., the increased rate of gaze shift to M2) in the Other/None context corroborates enhanced vicarious reinforcement during social decision-making.

Rewarding the other monkey without any opportunity to reward self is a uniquely vicarious form of reward. Such vicarious reinforcement may be driven by an intrinsic tendency to observe the experience of others to gather information, as can occur in foraging (Cheney and Seyfarth, 1990; Valone and Templeton, 2002). It is possible, however, that monkeys simply find feedback to their actions intrinsically rewarding. For instance, choosing to reward others in the Other/None context is the only option that results in a salient feedback that could serve as a secondary reinforcer or confirmation that a chosen action has resulted in a noticeable change in the environment. However, the preference to reward only self in the Self/Both context makes this possibility less likely (although the actor monkeys may have been less interested in the other monkeys due to receiving reward or the competitiveness evoked by this context), since choosing to reward both would also result in salient feedback. Furthermore, the absence of preference in the non-social control trials indicates that mere actions that result in fluid delivery are not sufficient to drive vicarious reinforcement, suggesting that the presence of a social agent is required. Notably, however, monkeys still showed high error rates (71%) in the conditioning trials when



the visual cues predicted reward to other monkey only. Interestingly, the error rates were much lower ($<13\%$) when monkeys confronted a choice between Other/None in the preference task. This is consistent with observations that rhesus macaques are much more motivated when making voluntary choices compared to making simple operant responses (Suzuki, 1999). Still, the atypically large error rates observed in the conditioning trials seems to be consistent with the competitiveness of rhesus macaques, and may highlight differences between humans and rhesus macaques (also see below).

In contrast, any of the two available options from the Self/Both context results in direct fluid reward. The preference to withhold reward from others in this particular context may reflect a potential diminishment of reward during simultaneous consumption, possibly due to the uncertainty of the quantity or quality of reward delivered to others. Reward withholding behavior may also arise from rhesus monkeys' natural competitive tendencies

(Anderson and Mason, 1978). For instance, from an ecological standpoint, sharing food with other individuals always reduces the amount of potential food available to oneself. Moreover, reduced rates of attending to M2 in the Self/Both context may further mitigate vicarious reinforcement during social decision-making. We observed a small but significant tendency to withhold less if actor monkeys were dominant to recipient monkeys, although our limited sample size and types preclude strong conclusions. This is consistent with a recent study in long-tailed macaques (*M. fascicularis*) showing that dominant macaques are more "prosocial" toward subordinates (Massen et al., 2010). Dominant monkeys might be more likely to engage in such positive other-regarding behaviors to sustain their rank and promote group cohesion, especially when there is no added cost, as in the Self/Both context. By extension, we would predict humans to choose to reward both individuals in an analogous monetary version of the Self/Both context, as long as the monetary reward was the same for both individuals and the amount of reward was undiminished by sharing (i.e., non-competitive situation). If the monkeys were clearly aware that they both always received the same amount of juice with an infinite amount of resources, they might also increase preferences to reward both monkeys. Alternatively, it is also plausible that the rhesus macaques, unlike humans, have a difficult time in ignoring their naturally competitive cognitive set.

It is critical to emphasize the dramatic differences in preferences between Self/Both and Other/None contexts. If the actor monkeys always found it valuable to reward the recipient monkey, then we would have expected the monkeys to prefer to reward both in the Self/Both context. Alternatively, if the monkeys always found rewards delivered to the other monkey to be aversive, perhaps due to perceived competition, then we would have expected the monkeys to prefer to reward none in the Other/None context. Instead, we observed a clean dissociation of preferences depending on social context, suggesting that different reward contingencies strongly influenced decisions. This is consistent with our findings that RTs, frequency of attention directed to the other monkey, and error rates were clearly different between choosing to reward both and choosing to only reward other. (Please also see our response above for situation-specific social behaviors in humans and monkeys.) The behavioral and neural mechanisms responsible for such context-dependent social decision-making would provide new insights into the social flexibility characterizing the behavior of macaques and other primates, including humans.

We hypothesize that vicarious experiences are processed as rewarding signals in the brain, and are mediated by neurons in homologous circuits governing social perception and reward learning in non-human primates and humans (Bandura and Rosenthal, 1966; Fehr and Camerer, 2007; Lohrenz et al., 2007; Lee, 2008; Hayden et al., 2009; Mobbs et al., 2009). One plausible mechanism is that the overlapping populations of neurons respond both to rewards to self and rewards to another individual. Such vicarious reward could motivate social interactions as well as underlie observational learning and mutualistic behaviors such as alliance formation, social grooming, and group cohesion (Fehr and Fischbacher, 2003; Takahashi et al., 2009). Modulation of vicarious

reward signals by social variables such as dominance or familiarity could further provide a mechanism promoting socially adaptive behavior toward specific individuals.

Observing rewarding events of others has been shown to systematically and effectively modulate neural activity in classic reward areas in humans, including ventral striatum, ventromedial prefrontal cortex, and anterior cingulate cortex (Mobbs et al., 2009; Lombardo et al., 2010). Moreover, the anterior cingulate cortex has been implicated in evaluating social information with respect to others (Takahashi et al., 2009). Dorsolateral and ventromedial prefrontal cortices in humans have been implicated in observing an action and observing reward outcome of others, respectively (Burke et al., 2010). Observational fear conditioning in mice depends on affective pain circuitry including anterior cingulate cortex (Jeon et al., 2010). Activation of these neural circuits by vicarious outcomes may be the neural substrate that ultimately promotes empathy and altruism, as well as observational learning.

REFERENCES

- Anderson, C. O., and Mason, W. A. (1978). Competitive social strategies in groups of deprived and experienced rhesus monkeys. *Dev. Psychobiol.* 11, 289–299.
- Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm glow giving. *Econ. J.* 100, 464–477.
- Bandura, A., and McDonald, F. J. (1963). Influence of social reinforcement and the behavior of models in shaping children's moral judgments. *J. Abnorm. Psychol.* 67, 274–281.
- Bandura, A., and Rosenthal, T. L. (1966). Vicarious classical conditioning as a function of arousal level. *J. Pers. Soc. Psychol.* 3, 54–62.
- Bandura, A., Ross, D., and Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *J. Abnorm. Psychol.* 67, 601–607.
- Berger, S. M. (1962). Conditioning through vicarious instigation. *Psychol. Rev.* 5, 450–466.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Burkart, J. M., Fehr, E., Efferson, C., and van Schaik, C. P. (2007). Other-regarding preferences in a non-human primate: common marmosets provision food altruistically. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19762–19766.
- Burke, C. J., Tobler, P. N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14431–14436.
- Byrne, R. M. (2002). Mental models and counterfactual thoughts about what might have been. *Trends Cogn. Sci.* 6, 426–431.
- Call, J., and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192.
- Cheney, D. L., and Seyfarth, R. M. (1990). The representation of social relations by monkeys. *Cognition* 37, 167–196.
- Crawford, M. P., and Spence, K. W. (1921). Observational learning of discrimination problems by chimpanzees. *J. Comp. Psychol.* 27, 133–147.
- de Waal, F. B., Leimgruber, K., and Greenberg, A. R. (2008). Giving is self-rewarding for monkeys. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13685–13689.
- Drea, C. M., and Wallen, K. (1999). Low-status monkeys “play dumb” when learning in mixed social groups. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12965–12969.
- Epstude, K., and Roese, N. J. (2008). The functional theory of counterfactual thinking. *Pers. Soc. Psychol. Rev.* 12, 168–192.
- Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427.
- Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791.
- Flombaum, J. I., and Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Curr. Biol.* 15, 447–452.
- Hawley, P. H. (1999). The ontogenesis of social dominance: a strategy-based evolutionary perspective. *Dev. Rev.* 19, 97–132.
- Hayden, B. Y., Pearson, J. M., and Platt, M. L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science* 324, 948–950.
- Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S.-Y., Rabah, D., Kinet, J.-P., and Shin, H.-S. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca²⁺ channels in ACC. *Nat. Neurosci.* 13, 482–488.
- Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1, 411–416.
- Kiehl, K. A. (2006). A cognitive neuroscience perspective on psychopathy: evidence for paralimbic system dysfunction. *Psychiatry Res.* 142, 107–128.
- Klein, J. T., Shepherd, S. V., and Platt, M. L. (2009). Social attention and the brain. *Curr. Biol.* 19, R958–R962.
- Lakshminarayanan, V. R., and Santos, L. R. (2008). Capuchin monkeys are sensitive to others' welfare. *Curr. Biol.* 18, R999–R1000.
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409.
- Lohrenz, T., McCabe, K., Camerer, C. F., and Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9493–9498.
- Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., and Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *J. Cogn. Neurosci.* 22, 1623–1635.
- Maestripieri, D. (1994). Mother–infant relationships in three species of macaques (*Macaca mulatta*, *M. nemestrina*, *M. arctoides*). II. The social environment. *Behaviour* 131, 97–113.
- Massen, J. J., van den Berg, L. M., Spruijt, B. M., and Sterck, E. H. (2010). Generous leaders and selfish underdogs: prosociality in despotic macaques. *PLoS ONE* 5, e9734. doi: 10.1371/journal.pone.0009734
- Masserman, J. H., Wechkin, S., and Terris, W. (1964). “Altruistic” behavior in rhesus monkeys. *Am. J. Psychiatry* 121, 584–585.
- Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., Schweizer, S., Frith, C. D., and Dalgleish, T. (2009). A key role for similarity in vicarious reward. *Science* 324, 900.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.
- Preston, S. D., and de Waal, F. B. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20; discussion 20–71.
- Rescorla, R. A., and Wagner, A. R. (1972). *A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement*. New York: Appleton-Century-Crofts.
- Roesch, M. R., and Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* 304, 307–310.
- Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.
- Silk, J. B., Brosnan, S. F., Vonk, J., Henrich, J., Povinelli, D. J., Richardson, A. S., Lambeth, S. P., Mascaró, J., and Schapiro, S. J. (2005). Chimpanzees are indifferent to the welfare of unrelated group members. *Nature* 437, 1357–1359.
- Subiaul, F., Cantlon, J. F., Holloway, R. L., and Terrace, H. S. (2004). Cognitive imitation in rhesus macaques. *Science* 305, 407–410.
- Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. London: MIT Press.
- Suzuki, S. (1999). Selection of forced- and free-choice by monkeys (*Macaca fascicularis*). *Percept. Mot. Skills* 88, 242–250.
- Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., and Okubo, Y.

- (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* 323, 937–939.
- Tomasello, M., Call, J., and Hare, B. (2003). Chimpanzees versus humans: it's not that simple. *Trends Cogn. Sci.* 7, 239–240.
- Tomasello, M., Davids-Dasilva, M., Camak, L., and Bard, K. (1987). Observational learning of tool use by young chimpanzees. *Hum. Evol.* 2, 175–183.
- Valone, T. J., and Templeton, J. J. (2002). Public information for the assessment of quality: a widespread social phenomenon. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 357, 1549–1557.
- Whiten, A., McGuidan, N., Marshall-Pescii, S., and Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 2417–2428.
- Zentall, T. R., and Levine, J. M. (1972). Observational learning and social facilitation in the rat. *Science* 178, 1220–1221.
- Zentall, T. R., Sutton, J. E., and Sherburne, L. M. (1996). True imitative learning in pigeons. *Psychol. Sci.* 7, 343–346.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 05 January 2011; accepted: 20 February 2011; published online: 03 March 2011.
- Citation: Chang SWC, Winecoff AA and Platt ML (2011) Vicarious reinforcement in rhesus macaques (*Macaca mulatta*). *Front. Neurosci.* 5:27. doi: 10.3389/fnins.2011.00027
- This article was submitted to *Frontiers in Decision Neuroscience*, a specialty of *Frontiers in Neuroscience*.
- Copyright © 2011 Chang, Winecoff and Platt. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.