

constant throughout this range remains to be determined (Gibbon et al., 1997).

On this foundation, investigations of the neural mechanisms for timing and recording in memory the durations of intervals may proceed.

C.R. Gallistel

Rutgers Center for Cognitive Science
Rutgers University
Piscataway, New Jersey 08854

Selected Reading

- Barnet, R.C., and Miller, R.R. (1996). *J. Exp. Psychol. Anim. Behav. Process.* 22, 279–296.
- Barnet, R.C., Cole, R.P., and Miller, R.R. (1997). *Anim. Learn. Behav.* 25, 221–233.
- Buonomano, D.V., and Karmarkar, U.R. (2002). *Neuroscientist* 8, 42–51.
- Colby, C.L., and Goldberg, M.E. (1999). *Annu. Rev. Neurosci.* 22, 319–349.
- Fantino, E., and Goldshmidt, J.N. (2000). *Psychol. Sci.* 11, 229–233.
- Gibbon, J. (1977). *Psychol. Rev.* 84, 279–335.
- Gibbon, J., and Church, R.M. (1981). *J. Exp. Psychol. Anim. Behav. Process.* 7, 87–107.
- Gibbon, J., Malapani, C., Dale, C.L., and Gallistel, C.R. (1997). *Curr. Opin. Neurobiol.* 7, 170–184.
- Killeen, P.R., and Weiss, N.A. (1987). *Psychol. Rev.* 94, 455–468.
- Leon, M.I., and Shadlen, M.N. (2003). *Neuron* 38, this issue, 317–327.
- Meck, W.H., and Church, R.M. (1982). *J. Exp. Psychol. Anim. Behav. Process.* 8, 226–243.
- Morre, D.J., Chueh, P.J., Pletcher, J., Tang, X., Wu, L.Y., and Morre, D.M. (2002). *Biochemistry* 41, 11941–11945.
- Pennartz, C.M.A., de Jeu, M.T.G., Bos, N.P.A., Schaap, J., and Geurtsen, A.M.S. (2002). *Nature* 416, 286–290.
- Roberts, S. (1982). *J. Exp. Psychol. Anim. Behav. Process.* 8, 2–22.
- Roberts, S., and Church, R.M. (1978). *J. Exp. Psychol. Anim. Behav. Process.* 4, 318–337.

Principles of Pleasure Prediction: Specifying the Neural Dynamics of Human Reward Learning

Accumulating evidence from nonhuman primates suggests that midbrain dopamine cells code reward prediction errors and that this signal subserves reward learning in dopamine-receiving brain structures. In this issue of *Neuron*, McClure et al. and O’Doherty et al. use event-related fMRI to provide some of the strongest evidence to date that the reward prediction error model of dopamine system activity applies equally well to human reward learning.

The smell of fresh coffee brewing in the morning, the sight of a bright red lobster on the dinner plate, the touch of a lover’s lips before a kiss—all of these sensory signals provide information to our nervous system that a rewarding experience is soon forthcoming. By learning

the association between rewards and the sensory events that reliably precede them, our nervous system can gain the ability to appropriately anticipate and prepare for rewarding experiences. This learning capability extends to events or objects that are themselves affectively neutral but which can take on affective significance because of their systematic co-occurrence with rewards. Such effects were first conclusively demonstrated in the famous experiments of Pavlov, which showed that dogs can be conditioned to salivate in response to a ringing bell that had in the past been paired with the appearance of a slab of meat. Unfortunately, in humans, reward-based learning can also lead to the development of maladaptive behavior patterns, such as that of the pathological gambler who continually gets lured in by the sights and sounds of the casino, or the heroin addict who feels irresistible cravings at the sight of a syringe. The quest to discover the critical mechanisms underlying reward-based learning has been a prominent theme of psychological and neuroscience research over the last century. Over the last decade, however, there has been a rapid acceleration of scientific progress in this area, based in large part from two parallel streams of research: neurophysiological studies of conditioning in nonhuman primates and noninvasive neuroimaging studies of reward processing in healthy humans. In the primate conditioning work, important discoveries have been made regarding dynamic activity patterns of dopamine-producing cells of the midbrain during the course of learning (Schultz, 1998). In the human neuroimaging studies, recent methodological advances have enabled increasingly sophisticated probes of the neural systems that underlie reward-based learning (e.g., O’Doherty et al., 2002). In this issue of *Neuron*, O’Doherty et al. (2003) and McClure et al. (2003) report new findings that increase the level of convergence between these two research streams. Their work demonstrates that key learning-related phenomena observed in the primate dopamine studies can also be profitably investigated in humans.

The O’Doherty et al. and McClure et al. studies take as their starting point the finding that midbrain dopamine cells show stimulus-locked phasic bursts in activity that become modulated as a function of learning (Figure 1). Specifically, in one set of conditioning studies (Ljungberg et al., 1992; Schultz et al., 1993), monkeys were given a cue (the conditioned stimulus or “CS+”) that was reliably followed by a few drops of palatable fruit juice (the unconditioned stimulus or “US”). Initially, dopamine cells responded with phasically increased activity when the US was given. Over time, the monkeys apparently learned to expect the US following the CS+. The burst of activity shifted backward in time, so that phasic dopamine responses occurred when the CS+ was presented but not when the US was subsequently given as expected. Finally, in trials where the US was unexpectedly omitted, the dopamine cell activity was transiently depressed at the expected time of US delivery. These results have been interpreted as suggesting that dopamine cells code a reward prediction error signal, in which phasic activity bursts signal a discrepancy between actual and expected future rewards. Importantly, this error signal may serve as a direct trigger for learning associations to reward-predicting cues. Dopamine cells

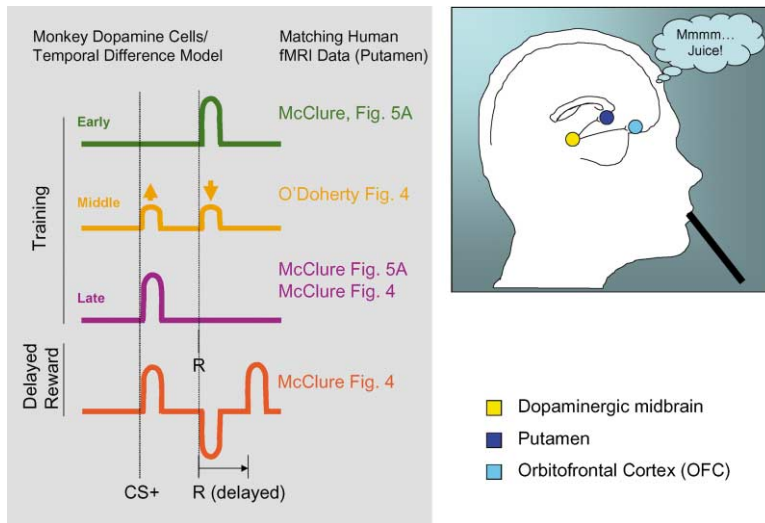


Figure 1. Neural Structures and Activity Dynamics in Reward Learning

Dopamine cells in the monkey midbrain respond with a burst of activity to unexpected reward (green). After pairing with a CS+, cells begin to respond to both the CS+ and subsequent reward R (gold). After training, cells respond only to the reward R (purple). If reward is unexpectedly delayed, the cells depress at the time of expected reward but burst later when reward is actually delivered (red). The McClure et al. and O'Doherty et al. studies found corresponding response properties in reward-related areas of the human brain, especially the putamen (dark blue) and OFC (light blue). After Schultz, 1998; Ljungberg et al., 1992; Hollerman and Schultz, 1998.

project to a number of different brain sites thought to be important for reward learning, such as the striatum and orbitofrontal cortex (OFC). Moreover, dopamine release has been found to be a key factor modulating synaptic plasticity (i.e., the neural expression of learning) at these sites (Reynolds et al., 2001). Finally, the reward prediction error hypothesis of dopamine activity fits strikingly well with previously developed formal models of reinforcement learning. In particular, one such learning algorithm, known as temporal differences (TD), makes use of phasic prediction error signals that shift in time across learning. Simulation studies have shown that the temporal behavioral of the TD error signal captures well the observed response profile of dopamine neurons during certain conditioning paradigms (Schultz et al., 1997). The putative link between formal learning theory and physiological mechanisms has provided a more precise and quantitative basis on which to generate new reward-learning experiments and evaluate their results.

The McClure et al. and O'Doherty et al. studies applied the clear experimental predictions derived from the primate work and the TD learning model to the examination of human reward learning. The primary question motivating both studies is to what extent is the neural system for reward learning in humans similar to that of the monkey? Event-related functional magnetic resonance imaging (fMRI) was used to indirectly measure brain activity to reward-related signals via monitoring of blood oxygen level dependent (BOLD) hemodynamic responses. A key feature of both studies was that human participants were tested in the identical types of conditioning paradigms used in the primate dopamine studies—passive (i.e., classical) conditioning procedures were applied with the same types of precisely timed juice rewards preceded by neutral visual cues. The McClure et al. study focused on the neural response to prediction errors caused by unexpected delays in the reward (e.g., 10 s after the cue instead of the expected 6 s). The O'Doherty et al. study included both omitted and unexpected rewards (occurring after CS-neutral cues) but primarily focused on the shift in neural activity

dynamics that were predicted to occur across trials with increased learning regarding the CS-US relationship. The primate work and TD model provided a clear set of predictions (see Figure 1): (1) in the initial phases of learning, transient activation increases should occur at the time of reward delivery, but in later phases, the activation increase should shift to the time of the CS+; and (2) unexpectedly delayed (or omitted) rewards should produce a transient decrease in activity at the expected time of reward delivery (negative prediction error) but a transient increase in activity when the unexpected reward is actually delivered (positive prediction error). A whole-brain search for regions that fit these particular patterns of activity dynamics identified the putamen (part of the striatum; both studies) and OFC (O'Doherty et al., 2003), confirming the central role of these dopaminergic-target structures in reward prediction learning. In addition, a strong feature of the O'Doherty experimental analysis was its direct utilization of the TD learning model to derive predictions regarding the temporal dynamics of reward-related brain activity and the shift in these dynamics with learning. Consequently, the O'Doherty study afforded a degree of rigor and quantitative precision to the evaluation of results which is a level above that typically achieved in neuroimaging studies. More generally, the approach of using computational models to generate regressors for neuroimaging data analysis seems especially promising, in terms of fostering greater interaction between theoretical, computational modeling, and experimental neuroimaging studies.

The McClure et al. and O'Doherty et al. studies should pave the way for a tighter integration and convergence between human and animal studies of reward conditioning, by establishing that comparable neural activity dynamics can be observed in humans and animals when exposed to the identical experimental paradigms. Thus, future studies can capitalize on the expected parallelism between animal and human experimental models, by exploiting new discoveries in one model system to drive further experimentation in the other. For example, a very recent report suggests that primate dopamine cells can

show sustained as well as phasic increases in cue-driven activity under conditions of maximal CS-US unpredictability (Fiorillo et al., 2003). It would be important to discover if such activity dynamics can be observed in humans as well. Conversely, studies of reward learning in humans admit a greater degree of experimental flexibility, given the ease and wide variety of contexts with which human subjects can be conditioned. Because fMRI methodology easily provides information about whole-brain activity, new reward-related brain areas might be discovered, and interactions among components of the reward network can be examined. In turn, these sorts of discoveries might provide the impetus for new animal studies by guiding the search for new brain regions to monitor with single-cell recordings (potentially under conditions where dopamine activity is simultaneously assessed). In terms of broader implications, the work discussed here might be fruitfully applied toward an understanding of the neural basis of normal individual differences (e.g., personality traits) and pathologies of reward learning (e.g., gambling and addiction), as these areas of study are still in their infancy (Breiter et al., 1997; Montague and Berns, 2002). For example, the selective probes of human reward conditioning developed in the McClure et al. and Doherty et al. studies might enable more precise quantification of abnormalities in the dynamic response profiles of different reward-related brain structures during conditioning.

The findings of McClure et al. and O'Doherty et al. also have implications for more theoretically oriented issues related to reward learning. Both studies examined classical (passive) rather than instrumental (response contingent) conditioning. Accumulating evidence suggests that the two forms of conditioning involve distinct mechanisms and potentially rely on non-overlapping neural substrates (Dayan and Balleine, 2002), yet this hypothesis has not been directly tested in humans. Likewise, alternative models to that of TD learning may provide a better account of dopamine system activity dynamics, including structures afferent to dopamine cells, particularly under conditions of premature rather than late reward delivery (Brown et al., 1999). Human imaging studies of such conditions could provide an important source of data to adjudicate between competing models. Regardless of the outcome of such future studies, the O'Doherty et al. and McClure et al. studies have raised the bar for the study of human reward processing by demonstrating how animal neurophysiological data and computational learning theory can be jointly exploited to probe reward-related brain structures in a more quantitatively and temporally precise manner.

Todd S. Braver and Joshua W. Brown
Department of Psychology
Washington University
St. Louis, MO 63139

Selected Reading

Breiter, H.C., Gollub, R.L., Weisskoff, R.M., Kennedy, D.N., Makris, N., Berke, J.D., Goodman, J.M., Kantor, H.L., Gastfriend, D.R., Riordan, J.P., et al. (1997). *Neuron* 19, 591–611.

Brown, J., Bullock, D., and Grossberg, S. (1999). *J. Neurosci.* 19, 10502–10511.

Dayan, P., and Balleine, B. (2002). *Neuron* 36, 285–298.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). *Science* 299, 1898–1902.

Hollerman, J.R., and Schultz, W. (1998). *Nat. Neurosci.* 1, 304–309.

Ljungberg, T., Apicella, P., and Schultz, W. (1992). *J. Neurophysiol.* 67, 145–163.

McClure, S.M., Berns, G.S., and Montague, P.R. (2003). *Neuron* 38, this issue, 339–346.

Montague, P., and Berns, G. (2002). *Neuron* 36, 265–284.

O'Doherty, J.P., Deichmann, R., Critchley, H.D., and Dolan, R.J. (2002). *Neuron* 33, 815–826.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). *Neuron* 38, this issue, 329–337.

Reynolds, J.N., Hyland, B.I., and Wickens, J.R. (2001). *Nature* 413, 67–70.

Schultz, W. (1998). *J. Neurophysiol.* 80, 1–27.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). *J. Neurosci.* 13, 900–913.

Schultz, W., Dayan, P., and Montague, P.R. (1997). *Science* 275, 1593–1599.