THE COGNITIVE NEUROSCIENCE OF MOTIVATION AND LEARNING

Nathaniel D. Daw New York University

Daphna Shohamy Columbia University

Recent advances in the cognitive neuroscience of motivation and learning have demonstrated a critical role for midbrain dopamine and its targets in reward prediction. Converging evidence suggests that midbrain dopamine neurons signal a reward prediction error, allowing an organism to predict, and to act to increase, the probability of reward in the future. This view has been highly successful in accounting for a wide range of reinforcement learning phenomena in animals and humans. However, while current theories of midbrain dopamine provide a good account of behavior known as habitual or stimulus-response learning, we review evidence suggesting that other neural and cognitive processes are involved in motivated, goal-directed behavior. We discuss how this distinction resembles the classic distinction in the cognitive neuroscience of memory between nondeclarative and declarative memory systems, and discuss common themes between mnemonic and motivational functions. Finally, we present data demonstrating links between mnemonic processes and reinforcement learning.

The past decade has seen a growth of interest in the cognitive neuroscience of motivation and reward. This is largely rooted in a series of neurophysiology studies of the response properties of dopamine-containing midbrain neurons in primates receiving reward (Schultz, 1998). The responses of these neurons were subsequently interpreted in terms of reinforcement learning, a computational framework for trial and error learning from reward (Houk, Adams, & Barto, 1995; Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997). Together with

Both authors contributed equally to this article. We are most grateful to Shanti Shanker for assistance with data collection, to Anthony Wagner for generously allowing us to conduct the experiment reported here in his laboratory, and to Alison Adcock, Lila Davachi, Peter Dayan, Mark Gluck, Mate Lengyel, Catherine Myers, Yael Niv, Shannon Tubridy, and Anthony Wagner for many fruitful discussions of the research reviewed here.

Correspondence concerning this article should be addressed to Daphna Shohamy, 1190 Amsterdam Ave., New York, NY 10027. E-mail: shohamy@paradox.psych.columbia.edu.

a large body of data suggesting an important functional role for dopamine and its targets in stimulus-response learning (Gabrieli, 1998; Knowlton, Mangels, & Squire, 1996; White, 1997), addiction (Everitt & Robbins, 2005; Hyman, Malenka, & Nestler, 2006), and movement (Albin, Young, & Penney, 1989; DeLong, 1990), these theories suggested the promise of a unifying account linking systems neuroscience with motivated behavior. However, despite their strengths, at a psychological level, these models are limited in their ability to capture many of the rich cognitive phenomena surrounding incentive motivation. Indeed, the theories closely parallel early behaviorist ideas, notably those of Thorndike and Hull.

Here we review this work, and discuss how these shortcomings can be remediated by situating it in a broader psychological and neural context that respects a more cognitive understanding of goal-directed behavior. In particular, we draw on operational distinctions elucidated in the animal behavioral literature in order to suggest that while reinforcement learning theories of dopamine may provide a good account of behavior known as habitual or stimulus-response learning, other neural systems are likely to be critically involved in goal-directed behavior. We discuss how this distinction resembles the classic distinction in the cognitive neuroscience of memory between nondeclarative and declarative memory systems, and review recent advances in the latter literature with an eye toward links between mnemonic and decision functions. Finally, we report a new experiment that probes the parallels between reinforcement learning and memory processes by demonstrating that representational processes characteristic of declarative memory emerge in the context of a reinforcement learning task.

DOPAMINE AND REINFORCEMENT LEARNING

The neuromodulator dopamine and its most prominent target, the striatum, have long been known to occupy a key point at the nexus of motivation and action. A major breakthrough in understanding the system came from the confluence of computational modeling and neurophysiology, when it was recognized that the activity of dopamine neurons recorded in primates receiving reward appears to carry a so-called "error signal" strikingly similar to those used for reinforcement learning in computer science (Houk et al., 1995; Montague et al., 1996; Schultz et al., 1997; Sutton and Barto, 1998).

Figure 1 illustrates in schematic form some key results supporting this interpretation (Fiorillo, Tobler, & Schultz, 2003; Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004). Here, a thirsty monkey irregularly receives drops of juice, which are signaled more or less reliably by prior visual cues. When reward is unexpected (when the cue preceding it is rarely reinforced, or when the reward arrives entirely uncued), dopamine neurons are phasically excited. However the response is not simply a report of reward: when reward is entirely expected on the basis of the preceding cue, the neurons do not respond to the reward. Moreover, when cues are partially predictive of reward, the strength of the phasic response to reward is modulated rather linearly by the *degree* to which the reward is expected. Finally, when reward is expected but fails to arrive, the neurons are briefly inhibited below their baseline firing rate. Altogether, dopamine neuronal firing relative to baseline appears to report the difference between observed and expected reward—a so-called reward prediction error.

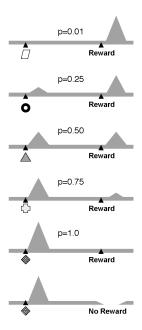


FIGURE 1. Schematic representation of midbrain dopamine neurons responding to reward that is probabilistically predicted by distinct visual cues; based on Fiorillo et al., 2003; Morris et al., 2004.

Such error signals can be used to update reward predictions and thereby incrementally improve them (e.g., in this case, to learn what reward probability is associated with each cue), a computational idea that in psychology goes back to the seminal conditioning theory of Rescorla and Wagner (1972). The temporaldifference learning algorithm (Sutton, 1988) that appears to explain the dopamine response generalizes that model essentially by chaining predictions backward in time. This is reflected in the fact that, in addition to responding to unpredicted rewards, the neurons also respond to reward-predictive cues. Hence, excitation trades off between the cue and the reward such that a fully predicted reward does not activate the neurons but instead the cue predicting it does. In the past several years, this model has been tested quite quantitatively (Bayer & Glimcher, 2005; Fiorillo et al., 2003; Morris et al., 2004; Satoh, Nakai, Sato, & Kimura, 2003) and has proven an extremely successful account of the dopamine response. That said, it has been suggested (e.g. by Horvitz, 2000, 2002; Redgrave & Gurney, 2006; Redgrave, Prescott, & Gurney, 1999) that dopamine neurons might not code prediction error for reward, specifically, but might rather report some more general signal, such as surprising events. On closer examination, however, the seemingly anomalous dopaminergic responses that in part motivated these suggestions can be accommodated in a reward prediction error model (Kakade & Dayan, 2002; Niv, Duff, & Davan, 2005).

These unit recordings mostly involve situations of prediction, with minimal action choice (though see Bayer & Glimcher, 2005; Morris, Nevet, Arkadir, Vaadia, & Bergman, 2006; Roesch, Calu, & Schoenbaum, 2007). Nevertheless, the strong

suggestion of the computational models is that the purpose of predicting rewards is to choose rewarding actions: that is, that the system supports instrumental conditioning. On a more causal level, pharmacological and lesion studies in animals also support the involvement of dopamine in both Pavlovian and instrumental conditioning (e.g., Faure, Haberland, Condé, & El Massiou, 2005; Parkinson et al., 2002). Interestingly, the complex behavioral pharmacology of dopamine has led a number of authors to suggest that dopamine subserves aspects of behavior involving performance rather than learning; for instance, motivation ("wanting" or "incentive slience") or behavioral vigor (Berridge, 2007; Robbins & Everitt, 2007; Salamone, 2007). Though different in their emphasis, these theories are, on our view, all essentially complementary to the viewpoint described here (Niv, Daw, Joel, & Dayan, 2007; McClure, Daw, & Montague, 2003; Niv, Joel, & Dayan, 2006).

Reinforcement learning models of dopamine have also had an important influence on the understanding of learning and motivation in human cognitive neuroscience. There is much evidence suggesting that similar midbrain dopamine mechanisms underlie a variety of reward-related behaviors in humans. Functional imaging (fMRI) has repeatedly demonstrated that the BOLD signal at striatal targets correlates with a temporal-difference prediction error, apparently reflecting dopaminergic innervation there (McClure, Berns, & Montague, 2003; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003). Initial studies of this sort were literal replications of the original primate designs; more recent work has extended these to various sophisticated decision situations (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006); the further fractionation of computational subunits (O'Doherty et al., 2004; Tanaka et al., 2004); the effect of dopaminergic medications on the neural signal and choice behavior (Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006); and prediction errors in uniquely human situations such as social interactions, and with a wide range of rewarding stimuli including money, verbal feedback, beauty, and juice (Aron et al., 2004; Delgado, Frank, & Phelps, 2005; Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; Haruno & Kawato, 2006; Haruno et al., 2004; King-Casas et al., 2005; McClure, Li, et al., 2004; O'Doherty, 2004; Tricomi, Delgado, & Fiez, 2004).

Further evidence for the involvement of midbrain dopamine in learning in humans comes from studies examining learning in individuals with Parkinson's disease. This disease causes a profound loss of dopamine-containing neurons in the midbrain, leading to dopamine depletion in the striatum (Graybiel, Hirsch, & Agid, 1990). While the most prominent symptoms of the disease are motoric, recent studies demonstrate that the disease also causes particular cognitive deficits in learning about reward (Frank, Seeberger, & O'Reilly, 2004; Gabrieli, 1998; Shohamy, Myers, Geghman, Sage, & Gluck, 2006; Shohamy, Myers, Grossman, Sage, & Gluck, 2005; Shohamy, Myers, Grossman, et al., 2004). In particular, Parkinson's disease specifically impairs incremental, feedback-based, stimulus-response learning, while other forms of learning remain intact (Gabrieli, 1998; Knowlton et al., 1996; Shohamy, Myers, Grossman, et al., 2004). These findings are among those supporting the suggestion that the striatum subserves a specialized system for habit or procedural learning (Gabrieli, 1998; Knowlton et al., 1996; White, 1997).

The neurophysiological data and computational models further suggest that the signals hypothesized to be important for learning are brief, tightly timed phasic events riding atop a baseline activity level. Since standard medications for Parkinson's disease apparently elevate dopaminergic tone more globally, they might ac-

tually mask or wash out such signals, and thereby *impair* feedback-based learning even while remediating grosser motor deficits. Several recent studies support this hypothesis, demonstrating that dopaminergic medication in Parkinson's patients can impair feedback-based stimulus-response learning (Cools, Barker, Sahakian, & Robbins, 2003; Frank et al., 2004; Shohamy et al., 2006). Along similar lines, Frank and colleagues showed that dopaminergic medication differentially affected learning from positive and negative feedback, as expected from their effects on different aspects of error signals in their computational models (Frank, 2005; Frank et al., 2004). Also in the modeling domain, Niv and colleagues have suggested an elaboration of the temporal-difference model in which overall dopaminergic tone controls behavioral vigor while the phasic signals drive learning, consistent with these sorts of dissociations between pharmacological effects on motor and cognitive symptoms (Niv et al., 2007).

In summary, the idea of a dopaminergic prediction-error system driving learning from reward has had extremely broad explanatory power, and appears to span a range from unit neurophysiology to motivated behavior in both humans and animals.

MOTIVATION AND GOAL-DIRECTED BEHAVIOR

We have discussed both physiological and behavioral evidence supporting the idea that dopamine is involved in reinforcement learning. However, a key challenge in interpreting neural manipulations such as lesions or psychopharmacology is the possibility that apparently intact behavior could be subserved by compensatory systems. Indeed, among the most important conclusions arising from the study of conditioning in animals and learning in humans is the finding that a single behavior (such as a lever press, or a choice response) can potentially arise from multiple processes that are both behaviorally and neurally dissociable (Cardinal, Parkinson, Hall, & Everitt, 2002; Shohamy, Myers, Kalanithi, & Gluck, 2008; Shohamy, Myers, Onlaor, & Gluck, 2004). In this section, we discuss how the dopaminergic reinforcement learning system fits into this bigger picture.

In animal studies, a particularly important dissociation has been drawn out by experiments probing the mnemonic representations underlying a particular action: specifically, whether an action is, or is not, driven by knowledge of the specific expected rewarding outcome ("goal," such as food) (for a fuller review see Dickinson & Balleine, 2002). Interestingly, experiments reveal that rats sometimes demonstrate knowledge of the outcome, but under other circumstances behave as though they are ignorant of it.

A typical such experiment relies on a reward devaluation test. First, a hungry rat is trained to lever press for food, and then a lever pressing test is conducted under circumstances when the animal does not want the food: for instance, after the animal is fed to satiety and will not eat the food, if given it. The critical question is whether the rat, after being fed, will stop pressing the lever during the test phase, or whether it will instead continue to press the lever, despite no longer being hungry. The test is conducted without food being provided so that a reduction of leverpressing in the test phase (relative to leverpressing for some outcome that is still desired) is attributable to the animal knowing which outcome is associated with the action.

In fact, under some circumstances a previously hungry rat reduces its lever pressing once fed to satiety. Under other circumstances, lever pressing-behavior persists unaffected even after the animal has been fed to satiety. When devaluation triggers a decrease in lever pressing, behavior demonstrably reflects knowledge of the associated goal; thus, such behavior has been defined as goal-directed (Adams, 1982; Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995). By contrast, when lever pressing persists even after devaluation of the outcome, such behavior is thought to arise from a "stimulus-response habit" stamped in by previous reinforcement rather than specific knowledge of the goal, and thus has been categorized as habitual rather than goal-directed (Adams, 1982; Dickinson et al., 1995). Of course, such devaluation insensitive behavior does not demonstrate that animals are literally ignorant of the action-outcome contingency, only that this information does not impact their decision to lever press. It is important to note that, because it excludes habits and requires that behavior demonstrably reflect goal knowledge, the definition of "goal directed" that we adopt here is stricter than the way the term is sometimes used in other areas of psychology.

A number of factors seem to impact the tradeoff between goal-directed and habitual behavior in devaluation studies. Notably, knowledge about goal identity appears to support behavior early in training, which is often devaluation sensitive, but behaviors often become devaluation insensitive following further training, evidently transitioning to habitual control (Adams, 1982; Dickinson et al., 1995).

Similar to habits, standard temporal-difference learning models of dopaminer-gic reinforcement learning do not learn or make use of any outcome knowledge: they work instead by learning only the overall desirability (generic "future value") of candidate actions (see Daw, Niv, & Dayan, 2005). For this reason, like early stimulus-response theories from behaviorist psychology, these theories would predict insensitivity to an outcome devaluation test, much like animals in the habitual stage of behavior. Notably, lesions to dorsolateral striatum and dopamine depletions of dorsal striatum render behavior persistently devaluation sensitive (i.e. they appear to disrupt habit formation and leave behavior perpetually goal directed; Faure et al., 2005; Yin, Knowlton, & Balleine, 2004).

On the whole, then, there is a striking correspondence between the psychological category of habits, the computational properties of the temporal-difference learning algorithm, and the physiology and functional anatomy of the dopaminergic system.

If theories of dopaminergic reinforcement learning correspond to the psychological category of habits, how can we extend this understanding to the separate category of goal-directed behavior and its neural substrates? Much less detail is known, either at the physiological or computational level, about goal-directed behavior (as defined operationally above). Speculation centers on the prefrontal cortex, which in human neuropsychology and imaging seems to be particularly implicated in goal representation, decision, and planning (e.g. Corbit & Balleine, 2003; Killcross & Coutureau, 2003; Owen, 1997; Valentin, Dickinson, & O'Doherty, 2007). In fact, lesions to a broad network of regions, including the prelimbic division of prefrontal cortex, disrupt goal-directed behavior in rats, leaving responding devaluation-insensitive, apparently habitual (Balleine & Dickinson, 1998; Corbit & Balleine, 2003; Killcross & Coutureau, 2003). Reflecting the breadth of this network, recent lesion results suggest that the goal-directed vs. habitual distinction may not map simply to the traditional prefrontal vs. striatal distinction, but

rather that each category of behavior may actually implicate a distinct corticostriatal "loop," each incorporating parts of both prefrontal cortex and striatum (Yin, Ostlund, Knowlton, Balleine, 2005; Killcross & Coutureau, 2003).

Computationally, goal-directed behavior seems to correspond well to a different category of reinforcement learning algorithms: so called *model-based* learning (Bertsekas & Tsitsiklis, 1996). Unlike temporal-difference learning, such algorithms plan actions by learning a representation—called a world model—of the contingencies in the task, including which actions lead to which outcomes. Like goal-directed behavior, then, model-based learning is rooted in knowledge about outcomes and is therefore capable of showing immediate, directed motivational sensitivity. For these reasons, variations on model-based learning have been suggested as candidates for reinforcement learning models of goal-directed behavior (Daw et al., 2005; Daw, Niv, & Dayan, 2006; Dayan & Balleine, 2002). However these models do not yet enjoy the same level of physiological and behavioral constraints as do the temporal-difference learning models.

To summarize, for all its successes, the most celebrated and better understood half of the cognitive neuroscience of reward learning is also the one least deserving the name "cognitive"—the stamping-in, under dopaminergic control, of stimulus-response habits. Such processes appear to be neurally, behaviorally, and computationally distinct from goal-directed action. This underlines the need for a similarly detailed understanding of goal-directed actions, both in themselves and in terms of how they interact with habitual influences.

Another fundamental open issue is whether similar dual processes underlie learning in humans. The work discussed in this section was largely developed in the rat, and efforts are only now under way to develop similar tasks in humans (Valentin et al., 2007). Although existing data do suggest a parallel between the neural mechanisms in humans and animals for the learning of habits (Frank et al., 2004; Gabrieli, 1998; Knowlton et al., 1996; Shohamy et al., 2006; Shohamy et al., 2005; Shohamy et al., 2004a), note that according to the strict definitions used in the animal studies, nearly all of the tasks used in human studies are actually *ambiguous* as to whether the tested behavior was goal-directed or habitual, since they did not probe outcome knowledge using a test such as devaluation (though see Valentin et al., 2007).

Where else, then, might we look for a deeper and more human-oriented understanding of goal-directed action? One suggestive insight is that goal-directed action is specifically defined here in terms of a representational or mnemonic demand—it is behavior dependent on knowledge about the identity of the expected reward. The domain of memory has long showcased a distinction between declarative and nondeclarative processes (Gabrieli, 1998; Knowlton et al., 1996; Poldrack et al., 2001; Squire, 1987). The former have been closely associated with rapidly formed, explicit representations and the medial temporal lobe (MTL); the latter with stimulus-response habits and with the striatum. More recently, declarative memory has also been shown to involve prefrontal cortex (for review, see Paller & Wagner, 2002; Simons & Spiers, 2003; Wagner, Koutstaal, & Schacter, 1999), which is also traditionally implicated in goal-directed action.

All of these considerations suggest parallels between goal-directed behavior and declarative memory. Indeed, it has long been suggested that the knowledge underlying goal-directed behavior is represented in declarative memory (Dickinson, 1980). In the rest of this article, therefore, we review the idea of multiple memory

systems with a particular eye toward drawing out implications for reward and motivation. Finally, we conclude with an experiment that explores one such analogy.

MULTIPLE MEMORY SYSTEMS

Decades of research in the cognitive neuroscience of learning and memory have led to the widely held view that memory is subserved by multiple independent cognitive and neural systems (Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001; Gabrieli, 1998; Robbins, 1996; Squire, 1987). At the broadest level, long-term memory is often separated into declarative and nondeclarative processes.

One relatively well explored declarative system is thought to support long-term memory for events or episodes—referred to as episodic memory (Gabrieli, 1998; Paller & Wagner, 2002; Squire, 1987; Wagner, Koutstaal, & Schacter, 1999). Episodic memories are formed rapidly (after even a single experience), and their representations are rich in contextual details, including representation of the relation between multiple arbitrarily associated stimuli. Episodic memories are also flexible, can be retrieved and accessed based on partial cues, and can be generalized to novel stimuli and contexts (Cohen, 1984; Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001).

Because these rich details and the relation between them are accessible for mnemonic retrieval, the subjective experience of episodic memory in humans is often explicit, and involves conscious awareness. It is worth noting, however, that there are no data to support the presumption that episodic representations are themselves conscious; in fact, a growing number of studies suggest that episodic representations can drive behavior without any evidence for conscious awareness, in both humans and animals (Barense et al., 2005; Chun & Phelps, 1999; Daselaar, Fleck, Prince, & Cabeza, 2006; Dusek & Eichenbaum, 1997; Griffiths, Dickinson, & Clayton, 1999; Lee et al., 2005; Preston & Gabrieli, 2008; Ryan & Cohen, 2004; Schnyer et al., 2006; for discussion see Ferbinteanu, Kennedy, & Shapiro, 2006).

Episodic memory is traditionally contrasted against the long-term memory for procedures or habits, as discussed above, a form of nondeclarative memory. This type of learning is characterized by incremental acquisition of stimulus-response associations over many experiences, and is thought to be stimulus-specific and inflexible (Eichenbaum & Cohen, 2001; Gabrieli, 1998; Knowlton et al., 1996; Robbins, 1996).

Converging evidence from studies in humans and animals indicates that episodic memory depends critically on the MTL, including the hippocampus and surrounding MTL cortices (Cohen & Squire, 1980; Eichenbaum & Cohen, 2001; Gabrieli, 1998; Knowlton et al., 1996; Paller & Wagner, 2002; Squire, 1987, 1992; Wagner et al., 1998). In humans, damage to the MTL impairs new episodic learning while sparing other learning processes (Cohen & Squire, 1980, 1981; Eichenbaum & Cohen, 2001; Myers et al., 2003; Squire, 1992). Similarly, in animals, damage to the MTL leads to impairments in rapid learning of arbitrary associations between co-occurring stimuli, while gradual stimulus-response learning remains intact (Eichenbaum, Stewart, & Morris, 1990; Squire, 1987, 1992). Recent fMRI data provide an even tighter link between MTL activity and episodic memory, demonstrating that the extent of MTL activity during learning predicts the formation of

successful episodic memories (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Kirchhoff, Wagner, Maril, & Stern, 2000; Otten, Henson, & Rugg, 2001; Paller & Wagner, 2002; Schacter & Wagner, 1999; Wagner et al., 1998).

Interestingly, then, the characteristics of MTL-based episodic memories resemble those of goal-directed actions (e.g., Dickinson et al., 1995): they are formed rapidly, support early learning, and are flexible. This suggests that the MTL may contribute to goal-directed behavior. Relatively few studies have directly examined the link between the MTL system and goal-directed action. However, some support for this hypothesis comes from lesion studies in animals demonstrating that hippocampal lesions disrupt one measure of goal-directed behavior in rats (Corbit & Balleine, 2000), albeit not the devaluation test described above. In rat maze navigation tasks, hippocampal lesions also disrupt a "spatial" strategy that predominates early in training (and which therefore may be analogous to goal-directed action). After further training, an apparently habitual "response" strategy predominates, which is instead sensitive to dorsal striatal damage (Packard, Hirsh, & White, 1989).

Further, the prefrontal cortex is traditionally associated with goal-directed behavior, and is also increasingly recognized for a role in episodic memory. Recent functional imaging studies demonstrate that the extent of activity in PFC during encoding of items is predictive of successful subsequent memory for those items, as observed in the MTL (Blumenfeld & Ranganath, 2006; Kirchhoff et al., 2000; Otten et al., 2001; Paller & Wagner, 2002; Simons & Spiers, 2003; Wagner et al., 1999; Wagner et al., 1998). PFC is thought to contribute to episodic memory by interacting with the MTL to control and guide mnemonic processes necessary for both successful encoding and retrieval of memories. Multiple PFC subregions are thought to differentially support distinct aspects of episodic memory, both during encoding and retrieval (Simons & Spiers, 2003; Wagner & Davachi, 2001; Dobbins, Rice, Wagner, & Schacter, 2003; Kahn, Davachi, & Wagner, 2004; for review see Paller & Wagner, 2002).

Precisely which subregions of PFC subserve goal-directed actions in humans is still somewhat unclear, but the dorsolateral PFC is a likely point of intersection between goal-directed and episodic processes. Dorsolateral PFC lesions are associated with impairments in planning and decision-making (perhaps related to core cognitive contributions also similar to those supporting episodic memory, such as executive, attentional, and working memory functions, e.g., Manes et al., 2002). It is unclear whether dorsolateral PFC, or instead more medial frontal territories in humans, is a human analogue of rat prelimbic PFC, where lesions abolish goal-directed responding (Balleine & Dickinson, 1998; Fuster, 1997). Medial frontal areas, ventromedial PFC and medial OFC, are also closely associated with reward and decision making (e.g., Bechara, Damasio, Tranel, Damasio, 1997; O'Doherty, 2004; Valentin et al., 2007), though these are sometimes viewed as more closely allied with dopaminergic/striatal habit systems (e.g., McClure, Berns, & Montague, 2004).

Another link between episodic memory and goal-directed behavior is suggested by recent intriguing findings demonstrating that a network involving both MTL and PFC may be involved in imagining episodes in the future (Addis, Wong, & Schacter, 2007; Buckner & Carroll, 2007; Hassabis, Kumaran, Vann, & Maguire, 2007; Schacter & Addis, 2007a, 2007b; Szpunar, Watson, & McDermott, 2007). Functional imaging studies reveal overlapping networks that are active in both

remembering and imagining tasks (Addis et al., 2007; Szpunar et al., 2007). MTL patients, known for their mnemonic difficulties, are also less famously impaired when asked to imagine and describe specific future events (Hassabis et al., 2007).

All this has notable resonance with goal-directed behavior, which depends on a representation of the specific reward expected, in the future, for the candidate action. Computational accounts of goal-directed behavior in terms of model-based reinforcement learning stress the evaluation of candidate courses of action by enumerating their anticipated future consequences (Daw et al., 2005). In their review, Buckner and Carroll stress that all these examples involve "self projection" (Buckner & Carroll, 2007); we would prefer to stress the more mundane point that planning and imagining draw on memories.

Finally, as mentioned, an additional characteristic of MTL-based episodic memory that has important parallels with goal-directed behavior is *representational flexibility*—the ability to retrieve, access, and generalize learned knowledge in novel contexts and settings (Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001). Habit learning, by contrast, is thought to result in the formation of relatively inflexible representations that are specific to the stimulus and context in which they were learned (Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001; Gabrieli, 1998; Myers et al., 2003; White, 1997).

This dissociation has been demonstrated in animals and humans using two-phase learning and transfer tasks to assess MTL and striatal contributions to learning and flexibility in a single task. In these studies, subjects first engage in incremental stimulus-response learning, then are probed to transfer, generalize, or reverse what they have learned to novel contexts, stimuli, or feedback. MTL damage specifically impairs the ability to flexibly transfer or generalize, without significantly impacting the ability to learn individual associations (Buckmaster, Eichenbaum, Amaral, Suzuki, & Rapp, 2004; Bunsey & Eichenbaum, 1996; Myers et al., 2003). Striatal damage results in the opposite pattern: slow learning, but intact generalization (Myers et al., 2003; Shohamy et al., 2006).

Recent fMRI data from healthy individuals further demonstrate that increased hippocampal activity during learning relates successful flexible use of acquired knowledge (Foerde, Knowlton, & Poldrack, 2006; Shohamy & Wagner, 2007), while the opposite relation was found between activity in the striatum and knowledge transfer (Foerde et al., 2006; Heckers et al., 2004; Preston et al., 2004). The analogy with goal-directed behavior goes back at least to Dickinson (1980), who pointed out that the act of combining action-outcome and outcome value information to adjust behavior following outcome devaluation is itself a hallmark example of the representational flexibility that characterizes declarative knowledge.

Interestingly, although there is a great deal of evidence dissociating both goal directed and episodic systems from habits, many issues remain puzzling regarding the nature of the relationship between these systems. First, in the domain of memory, there is some evidence that the striatum and MTL may not be independent, but rather may competitively interact during learning (Packard et al., 1989; Poldrack et al., 2001; Poldrack & Packard, 2003). Both competitive and cooperative interactions have variously been proposed between goal-directed and habitual systems in reinforcement learning (Daw, Courville, & Touretzky, 2006; Daw et al., 2005).

Second, and perhaps related, it is important to note that the neural substrates for these systems are all substantially intertwined, suggesting richer interactions than so far understood. As already noted, despite proposed functional divisions between PFC and striatum, the PFC and striatum are famously interconnected via corticostriatal "loops," and MTL connects with both territories (Alexander, DeLong, & Strick, 1986; Fuster, 1997; Goldman-Rakic, Selemon, & Schwartz, 1984; Haber, 2003; Suzuki & Amaral, 2004). Not surprisingly then, more recent work on the functional anatomy of conditioning suggests that goal-directed and habitual systems each involve both prefrontal and striatal contributions (Killcross & Coutureau, 2003; Yin et al., 2004; Yin et al., 2005). Moreover, all of these regions (and not just striatum) are substantially innervated by dopamine. Indeed, intriguing new data indicate that midbrain dopamine, motivation, and reward all modulate MTL-based episodic memory, as well as more traditional reinforcement learning (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Wittmann et al., 2005; Shohamy & Wagner, 2007).

MEMORY SYSTEMS: SUMMARY AND IMPLICATIONS FOR REINFORCEMENT LEARNING

The studies reviewed here suggest that behavior is guided by multiple forms of memory, subserved by different neural systems. These memory systems parallel in their anatomical and functional characteristics systems that have been more or less independently described in the context of motivation and goal-directed behavior. These two literatures are complementary in that the dopaminergic/habitual system has been more deeply studied in reward learning while more is known about the mnemonic functions of the episodic/MTL system than about the goal-directed action system with which we identify it.

So far, we have discussed theoretical, conceptual and anatomical links between the characteristics of systems described for different mechanisms of motivated action, and between memory systems underlying different mnemonic processes. Although the parallels between these systems are compelling on the surface, fundamental open questions remain regarding how the memory systems approach informs our understanding of goal-directed motivated behavior.

The parallel neural and cognitive mechanisms implicated in both declarative memory and goal-oriented action suggest that mnemonic processes may contribute to motivated choice behavior. Key uncertainty concerns the nature of the relationship between memory and choice processes and their respective neural substrates (for example, are they essentially one and the same, or are they separate, with a goal-directed action system making use of an otherwise independent mnemonic system?). Further, at present very little is known regarding the behavioral implications of mnemonic processes for motivated choice behavior, as these two aspects of behavior have typically been assessed independently. In the next section, we present a preliminary attempt to bridge this gap by directly examining mnemonic representational changes that occur during reinforcement learning.

BRIDGING BETWEEN MOTIVATIONAL AND MNEMONIC ASPECTS OF LEARNING AND CHOICE: SIMULTANEOUS REPRESENTATIONAL AND REINFORCEMENT LEARNING

The flexible transfer of knowledge has been explored in both animals and humans using the "acquired equivalence" paradigm (Bonardi, Graham, Hull, & Mitchell, 2005; Grice & Davis, 1958; Hall, Ray, & Banardi, 1993; Myers et al., 2003). In acquired equivalence, prior training to treat two stimuli as equivalent increases later generalization between them—even if those stimuli are superficially very dissimilar. In a standard protocol, subjects first learn that two stimuli (such as two faces, S_1 and S_2), are both associated with the same outcome (such as a specific visual scene; O₁), while two other stimuli, S₂ and S₄, are both associated with a different outcome (a different visual scene; O₂), making two pairs of stimuli that are equivalent to one another in respect to their prediction of an outcome (S_1-S_2, S_2-S_4) . Next, subjects learn that S₁ is also associated with a different outcome O₃ and S₂ with O₄. Subsequently, when probed as to whether S₃ predicts O₃ or O₄, subjects tend to respond "O₃" despite having no prior experience with an S₃-O₃ pairing. This suggests that subjects learn equivalencies between stimuli with identical outcomes such that they are able to flexibly generalize their knowledge about outcomes from one stimulus to another.

Converging evidence suggests that acquired equivalence, as well as other forms of flexible transfer, depends on the MTL but not on the striatum (Myers & Gluck, 1996; Myers et al., 2003; Shohamy et al., 2006). A common interpretation of this phenomenon is that it reflects changes at the level of representation of the stimuli: S_1 and S_2 (and S_2 and S_4) are coded as more similar to one another in light of their similar outcomes, causing knowledge subsequently learned about either to generalize automatically (Gluck & Myers, 1993; Myers & Gluck, 1996).

Here, we developed a learning and transfer task that adapts the acquired equivalence paradigm to a reinforcement learning context. Subjects first engaged in a reward-based choice task typical of reinforcement learning: subjects chose among four face stimuli on each trial $(S_1 - S_4)$, each of which was associated with a different probability of monetary reward. The reward probabilities were gradually and randomly changed, which drove dynamic shifts in subjects' preferences since they tend to choose options that are more often rewarded.

Unbeknownst to the subjects, the four stimuli were split into two pairs (S_1 & S_3 ; S_2 & S_4) within each of which the dynamic reinforcement probabilities were yoked. This gives rise to correlated reinforcement histories, masked somewhat by randomness in the reward delivery and by the subjects' own uneven sampling of the options. (Note that subjects are only informed about reward for the option they chose on a particular trial: in order to discover the current value of an option, they must therefore choose it.)

We hypothesized that if relational episodic processes are involved during reinforcement learning, then this manipulation should result in acquired equivalence for the stimulus pairs that share a common underlying probability of reinforcement. To test this hypothesis, we subsequently trained subjects with new reward probabilities for one of each stimulus pair (S_1, S_2) ; finally, we probed subjects' choices for the paired stimuli on which they had not been retrained (S_3, S_4) . If reinforcement-based choice is modulated by relational mnemonic processes, subjects

should shift their preferences about S_3 and S_4 to reflect the retraining of S_1 and S_2 . If reinforcement-based choice is independent of these processes—and instead is driven solely by stimulus-specific habits—then subjects would be expected to maintain the same response to S_2 and S_4 as prior to the retraining with S_1 and S_2 . We additionally hypothesized that any generalization between equivalent options would also be visible during the initial choice task, in the ongoing process of reward-driven preference adjustment. Here, if relational learning impacts reinforcement learning, then feedback about S_1 , for example, should also drive preferences about its paired stimulus, S_3 .

From a reinforcement learning perspective, the payoff structure involves hidden higher-order structure in the form of correlations between the stimuli's values. Simple, first-order reinforcement learning methods of the sort associated with the mesostriatal dopamine system would be blind to this structure, and would not be expected to show any transfer between equivalent stimuli. More sophisticated learning mechanisms—such as "model learning" about the higher-order task structure—would be required to exhibit transfer. These, as discussed, are closely associated with goal-directed action.

METHODS

PARTICIPANTS

Data are reported from 19 healthy adults (10 females; ages 18-24 yrs); all were right handed, native English speakers. Data were lost from two additional participants due to software problems. All participants received \$10/hr for participation, with the experiment lasting approximately 1 hr. Informed written consent was obtained from all participants in accordance with procedures approved by the institutional review board at Stanford University.

TASK

The experiment consisted of three phases: Training, retraining, and transfer test. Sample screen events from each phase are shown in Figure 2.

The initial training phase (Phase 1) consisted of 400 trials. In each trial, four pictures of faces were presented on a computer screen and subjects had 3 seconds to choose between them using the keyboard. The face pictures were taken from the Stanford Face Database. The same four pictures were presented on each trial, and their locations were pseudorandomized. Three-quarter seconds after a choice was entered, subjects received feedback (a \$.25 reward, or none). This notification remained on the screen for 1 second; the screen was then blanked for a two-second intertrial interval and the next trial followed.

Subjects were instructed that each face was associated with a different probability of reward, that these probabilities could change slowly, and that their goal was to find the most rewarding face at a given time and choose it in order to earn the

Phase 1: Training

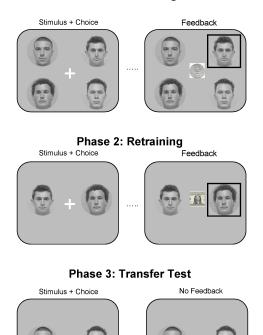


FIGURE 2. Sample screen events from learning and transfer phases of a novel acquired equivalence paradigm. See *Methods* for description of task phases.

most money. They were also instructed that rewards were tied to the face identity, and not the face position.

Rewards were probabilistically associated with each face. Unbeknownst to the subjects, the faces were grouped into equivalent pairs (here referred to as faces $S_1 \& S_3$ and $S_2 \& S_4$). The chance of reward on choosing S_1 or S_3 (and similarly S_2 or S_4) was the same at any particular trial. The reward probability for each pair of faces changed over time, however, diffusing between 25% and 75% according to independent Gaussian random walks with reflecting boundary conditions. Two instantiations of the random walks were used (i.e., two pairs of probability sequences, one illustrated in Figure 3), counterbalanced between subjects, and each selected so that the options' probabilities were nearly equal at the end of the training phase (to minimize preferences carried over into the transfer test). Also counterbalanced were the mapping of face pictures to numbers, and the mapping between the face pairs and the random walks.

The retraining phase (Phase 2) consisted of 40 trials that were the same as phase 1, except that subjects chose only between two faces (S_1 and S_2), and rewards, when they arrived, were \$1 rather than \$.25 (to promote a greater shift in context between the training and retraining phases). In this phase, reward probabilities for S_1 were constantly 20% and for S_2 , constantly 80%.

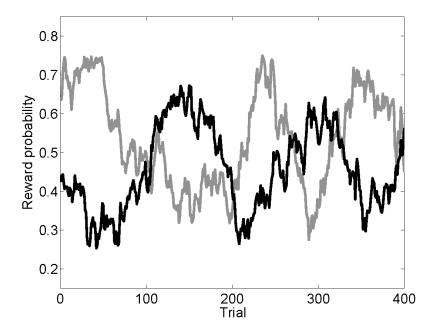


FIGURE 3. Example of probability that a choice of either face pair will be reinforced, as a function of trial number. Reinforcement probabilities change according to a random walk, as detailed in *Methods*.

After training, subjects were tested for transfer (Phase 3). Subjects were instructed that they would make additional choices between pairs of faces, with \$1 for each reward, but would not be informed of the reward until the end. All possible pairs of face combinations were presented 5 times (interleaved), with the critical probe trial (faces S_3 vs S_4) interleaved an extra 5 times, for a total of 35 trials.

At the end of the experiment, subjects were informed how many times they won in the last phase, and paid (as they had been previously informed) one-fifth of the total money earned across the three phases of the experiment. They then answered a series of questions assessing their strategies during learning and were debriefed.

Analysis: Learning and Transfer Performance. Transfer was assessed according to the fraction of choices (out of 10) of face S_4 over S_3 in phase 3 of the experiment. In this case, preference for S_4 over S_3 , for instance, was taken to be the number of choices of S_4 during that period divided by the number of choices of either S_3 or S_4 . The fraction of choices of face S_2 over S_1 was also assessed for comparison. Baseline preferences were assessed between faces S_4 and S_3 (and S_2 and S_1) for the end of the training phase (phase 1), using the last 10 choices from the phase (since preferences shifted rapidly, following fluctuations in the payoffs). Since they arose from counts over small numbers of trials, these measures were quantized and also bounded; they were also often highly skewed (since many subjects tended to make consistent choices over repeated trials, particularly during phases 2 and 3). For all these reasons, these measures are inappropriate for Gaussian statistics, and we accordingly compared them using nonparametric sign tests.

Model-Based Analysis of Choices. Phase 1 data were further analyzed by fitting a simple Q-learning model to the training phase choice sequences (Sutton & Barto, 1998). According to this, subjects assigned to each face a value $V_1 \dots V_4$ according to previous experienced rewards. These were assumed to be learned by a delta rule: if option c was chosen and reward r (1 or 0) received, then V_c was updated according to $V_c \leftarrow V_c + \eta(r - V_c)$. Here, the free parameter η controls the learning rate. Further, given value estimates on a particular trial, they were assumed to choose randomly between the options with probabilities $P_1 \dots P_4$ according to a softmax distribution: $P_c = \exp(\beta V_c)$, (Daw, O'Doherty, et al., 2006). The free parameter β controls the exclusivity with which choices are focused on the highest-valued option.

For each subject, parameters η and β were chosen using a gradient search to maximize the likelihood of the subject's observed choice sequence, conditioned on the rewards received. (That is, the product over trials of P_c for the chosen c using values learned by the model from the previously delivered rewards.)

In order to search for indications of acquired equivalence during the training phase, this model was compared to an elaborated model, in which feedback from choices of a face (say, S_1 or S_2) was also applied to learn about its equivalent partner (S_3 or S_4 , respectively). Of course, equivalence effects would be expected to be less than complete and to develop over time, as subjects gradually learned the equivalence. Absent a formal account of such higher-order learning, we consider a simplified model in which feedback about a face impacts its partner to a degree that is constant over the whole training phase. (Note that this will tend to underestimate any observed equivalence effect, by "averaging" over parts of the task before the equivalence could have been learned). In particular, if c was chosen, with partner p, then in addition to updating V_c as above, V_p was also updated according to $V_p \leftarrow V_p + \sigma(r - V_p)$ with a second free learning rate parameter σ . As a control, the same model was fit twice more, with the partner replaced by each of the noncorrelated faces (for face S_1 , S_2 and S_4).

Models were compared according to the likelihood they assigned to the training data, penalized for the additional parameter using the Bayesian Information Criterion (BIC; Schwartz, 1978). For each model, BIC scores were summed over subjects.

RESULTS

Two subjects were found to be performing a simple win-stay-lose-shift rule during the training phase (following it strictly for more than 95% of all choices). Since such an explicit rule-driven strategy may be qualitatively different from the more incremental learning apparently exhibited by the other subjects, we considered these subjects separately and do not include them in the group analyses presented below. Interestingly, unlike the remainder of the group, neither of these subjects shows indications of an acquired equivalence effect under any of the measures presented below. Nevertheless, if these subjects are included in the group analyses, the conclusions adduced below remain substantially the same. (In this case, one statistical test, noted below, fails to achieve significance; the remaining results do not change qualitatively.)

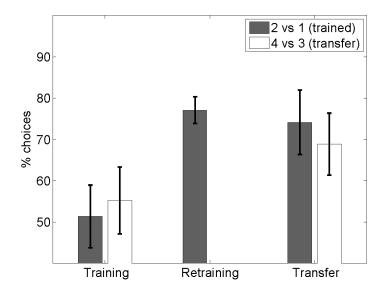


FIGURE 4. Mean (+/- 1SEM) percent choices for face S_2 vs S_1 (trained pair) and S_4 vs S_3 (transfer probe pair) in three experimental phases. At end of Phase 1 training ("initial") there are no systematic preferences between either of the two pairs of stimuli; a preference for S_2 over S_1 develops during retraining (Phase 2); this persists and transfers to preference for S_3 over S_4 in the final transfer test.

LEARNING AND TRANSFER PERFORMANCE

Figure 4 displays stimulus choices over the three phases of the experiment. As shown, at the end of the initial training phase (Phase 1), the population exhibited no preference for face S_2 over S_1 (median 50% choices of 2, sign test P=1 against null hypothesis of 50%), or for face S_4 over S_3 (median 60% choices of 4, P> .3). In the retraining phase (Phase 2), they developed a strong preference for face S_2 over S_1 (median 83% choices, P< 5e-5), and this preference carried over into the probe phase (median 80% choices, P< .05). Critically, in the probe phase (Phase 3), the population displayed a preference for face S_4 over face S_3 (median 70% choices, P< .05; this test does not attain significance when the two win-stay-lose-shift subjects are included), mirroring the trained preference for face S_3 over S_4 .

The acquired equivalence effect can also be observed in a within-subject comparison, by comparing each subject's preference for 4 over 3 in the transfer phase to her corresponding score at the end of the training phase. Here the median difference in preferences is 23% (P < .05 on a paired sign test).

We did not formally attempt to assess to what extent task learning was implicit or explicit. However, strikingly, despite clear acquired equivalence effects, when subsequently asked about what strategies they followed during the experiment, none of the subjects reported having noticed that stimuli's reward probabilities were yoked to each other during training. Instead, subjects reported a general attempt to discover the single most "lucky" face, and to avoid the "unlucky" ones. Subjects' inability to verbalize their use of equivalencies in driving their choices

is not surprising given the noisy stochastic distribution of reinforcement across the choices, and is consistent with other reports in humans of relational, flexible knowledge (typically MTL dependent) driving choices without conscious awareness (Daselaar et al., 2006; Lee et al., 2005; Preston & Gabrieli, 2008; Ryan & Cohen, 2004; Schnyer et al., 2006; Walther, 2002).

MODEL-BASED ANALYSIS OF CHOICES

To search for indications of acquired equivalence during the initial training phase, we also examined how choices shift trial-by-trial, and asked whether feedback about an individual stimulus also impacts subsequent choice of its equivalent partner. Such adjustments were examined by fitting a computational model of trial-by-trial choices to the raw data, as described in the Methods section. In brief, the simple reinforcement learning model assumes that subjects use rewards to estimate a value for each stimulus, and choose accordingly. We compared the fit of four models to subjects' trial-by-trial choices (see Methods): a standard baseline, which assumes subjects learn only about the stimuli they choose, an acquired equivalence augmentation in which experience about a stimulus additionally updates its partner's value (to a degree determined by a second learning rate parameter), and two control models, in which feedback about a stimulus instead updates either of the non-equivalent stimuli.

The free parameters of the models were fit per-subject to maximize the observed likelihood of her actual choice sequence. Small positive learning rates were found for feedback about a face affecting its equivalent partner (median 9% of the size of the learning rate for the chosen face), consistent with an acquired equivalence effect. These were only insignificantly larger than the median learning rates fit for nonequivalent faces in control models (each 2%, p > .15 on paired sign tests). However, only in the case of the equivalent faces were the additional parameters justified by correspondingly improved fit to the data, as assessed using the Bayesian Information Criterion score (Figure 5). As shown, the best fitting model was that containing learning about the equivalent faces, by a margin constituting very strong evidence against the other models according to the conventions suggested by (Kass & Raftery, 1995). Penalized for the additional parameters, the control models that learned about nonequivalent faces fit the data slightly worse than the base model. (It should be noted that these scores aggregate over the subjects; viewed individually, not every subject's fits justify learning about the equivalent face nor reject learning about the control faces).

On the whole, these findings suggest that an acquired equivalence-like effect can be observed even during the initial training phase.

DISCUSSION

The results of this experiment demonstrate, using a number of different measures, a robust acquired equivalence effect in the context of a trial-and-error reinforcement learning task in humans. Our methods, and our results, blend aspects from both the episodic and the reinforcement learning literatures. We show that flexible, relational representations—traditionally considered hallmarks of an episodic

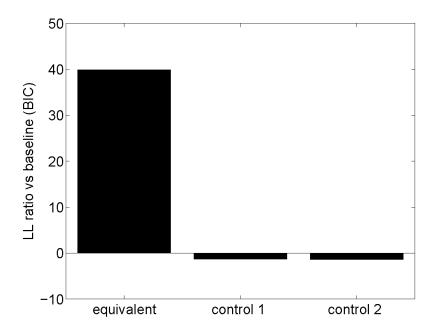


FIGURE 5. Penalized log likelihood ratios (difference in -BIC/2, larger is better) for three models of trial by trial choices versus a baseline model. Compared to the baseline model (which learns only about the chosen face), the model that additionally learns about the face equivalent to the chosen one ("equivalent") is strongly favored. The two control models, which each learn about one of the nonequivalent faces, fit the data no better than baseline.

memory system and associated with the MTL—develop during the course of reinforcement learning and guide subjects' choices. Specifically, during initial learning, subjects' choices of each individual stimulus appeared to be impacted by feedback received about its equivalent partner.

Furthermore, when probed in the final transfer phase, subjects' choices on the novel S_3 - S_4 pairing reflected their recent reinforcement experiences with two other, distinct, stimuli (S_1 - S_2), rather than their prior experience with S_3 and S_4 themselves. Such relational transfer would not be expected in simple stimulus-response forms of "habitual" reinforcement learning of the sort traditionally associated with striatum.

Acquired equivalence has previously been observed both in animals (using classical conditioning) and in humans using paradigms where equivalencies are set up using distinct stimulus-outcome pairings (Bonardi, Graham, Hall, & Mitchell, 2005; Grice & Davis, 1958, 1960; Myers et al., 2003). Prior neuropsychological data indicate that damage to the striatum slows initial associative learning, but does not impair generalization, while damage to the hippocampus impairs generalization, but spares initial learning (Myers et al., 2003; Shohamy et al., 2006). These lesion results emphasize a dissociation between learning (dependent on the striatum) and generalization (dependent on the hippocampus).

By contrast, the present findings demonstrate generalization between equivalent items even during initial learning. This result indicates that relational mechanisms contributed to reinforcement learning in this task. This accords with a major

point of the foregoing review, that trial-and-error reinforcement learning is not exclusively the domain of mesostriatal habit mechanisms but can also implicate a goal-directed system that functionally and anatomically mirrors that for episodic memory. Consistent with this interpretation, recent data from functional imaging (fMRI) demonstrate that hippocampal activity *during learning* predicts subsequent generalization in the transfer phase (Shohamy & Wagner, 2007).

From a reinforcement learning perspective, the results demonstrate that subjects exploited higher-order correlational structure in the reinforcement contingencies. This extends a recent finding (Hampton, Bossaerts, & O'Doherty, 2006) that subjects are able to make use of similar higher-order structure when instructed about it. Such learning is beyond the capacity of standard temporal-difference learning theories of striatal reinforcement learning. Interestingly, identifying or modeling these higher-order structures is computationally closely related to the sort of modeling (of action-outcome contingencies and outcome values) thought to underlie goal-directed action (Daw et al., 2005). In this sense, while we did not use a reward devaluation challenge to test for goal-directedness as in the animal conditioning literature, the acquired equivalence test for the impact of relational learning on choices plays a similar role. It will be important, in future, to compare these two assays in order better to understand to what extent the processes examined here correspond to the goal-directed versus habitual distinction from animal conditioning. If relational sensitivity reflects engagement of a goal-directed system, we would expect it also to predict sensitivity to devaluation.

One possible interpretation, therefore, is that choices in the present task originate wholly from a goal-directed reinforcement learning system, which (particularly since we have argued to identify it with MTL episodic memory processes) would be expected to incorporate the relevant relational structure. If this is true, then the analogy with the animal literature (in which behavior transitions to habitual with overtraining) suggests that relational information should have *less* influence following *more* training: i.e., the acquired equivalence effects should diminish.

Alternatively, and perhaps more intriguingly, the behavior observed here may arise from joint contributions of both systems. Much current debate focuses on the possibility, and the possible nature, of an interaction between nondeclarative and declarative learning, as well as between habitual and goal oriented behavior (Foerde et al., 2006; Poldrack et al., 2001; Poldrack & Packard, 2003). The present findings are consistent with the possibility that the choices observed here arise from a habitual, stimulus-response system, but are imbued with more sophisticated transfer capability by virtue of a cooperating representational process that encodes relational regularities between stimuli. By this account, equivalence in the reward probability of items leads to the development of similar representations for them, so that even simple stimulus-response learning about one will automatically transfer to the other.

A number of models of both acquired equivalence and reinforcement learning work just this way (Daw, Courville, et al. 2006; Gluck & Myers, 1993). Such representations can be derived from a learned model of the task structure (Daw, Courville, et al., 2006), or through other sorts of relational stimulus-stimulus learning. If this interpretation is correct, then the present experiment exercises both relational, episodic representation learning and habitual reinforcement learning processes, contributing cooperatively. This is an intriguing possibility given that,

as discussed in the present review, these processes are often viewed as separate and even competing. The further development of tasks such as this one should allow a careful analysis of these different components, how they are modulated by motivation and desirability, and how they change over time.

Interestingly, similar interactions between apparently different sorts of learning have also been reported in behavioral studies in the domain of social psychology, where two-system accounts of behavior at least superficially reminiscent to those discussed here are also popular (e.g., Lieberman, 2003; Strack & Deutsch, 2004). First, in an experiment with some resemblance to our own, (Walther, 2002) demonstrates a "spreading activation" effect of prior stimulus-stimulus associative conditioning on subsequent evaluative conditioning. In particular, when two stimuli (people's faces) co-occur, subsequent development of attitudes toward one person "spreads" to the other (Walther, 2002). As with acquired equivalence, this effect seems to occur independent of awareness in the context of implicit evaluative conditioning, but, notably, is demonstrated using sensory preconditioning—a conditioning procedure that is thought to depend on the hippocampus (Gluck & Myers, 1993; Port & Patterson, 1984). Conversely, in another experiment, implicit priming effects called nonconscious goal pursuit, which had been previously assumed to arise within a habitual-like associative system (Bargh, 1990; Eitam, Hassin, & Schul, 2008), have recently been shown to impact more explicit, presumably goal-directed, subsequent learning in a novel task (Bargh, 1990; Eitam et al., 2008). The demonstration of apparently cross-system learning across species, paradigms, and domains suggests parallels that deserve much further investigation.

CONCLUSIONS

Recent interest in the cognitive neuroscience of motivation and reward has focused on the role of midbrain dopamine neurons in reward prediction. This work has demonstrated that reward-prediction models of midbrain dopamine neurons successfully account for a wide—but also limited—range of motivated behaviors; specifically, those that underlie habitual, stimulus-response learning. A relatively less well-understood system appears to support what is referred to in behavioral psychology as goal-directed behavior. Here, we have outlined the neural and psychological substrates for these different aspects of behavior, and taken particular note of parallels between these systems and those involved in memory. We have further reported results from an experiment in which subjects simultaneously exercised mnemonic relational learning and reinforcement learning. The results, by design, tend to blur the sharp dichotomies between systems in these fields.

In the end, it should come as no surprise that decision making draws heavily on memory, and also that both rely on other shared cognitive capacities. That not-withstanding, the literatures have substantive parallels, complementary strengths, and, ultimately, great similarity in terms of the major open issues. These, in short, are issues about modularity. The great success of cognitive neuroscience, including in these areas, has been the fractionation of function; much remains to be understood about how these fractions reassemble into behavior. For instance, do relational and reinforcement learning processes cooperate or compete? How are

the various sorts of information underlying a goal-directed decision distributed throughout the brain? To what extent is this different from other sorts of declarative information? Such integrative questions represent the next steps for development in both areas.

REFERENCES

- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, 34B, 77-98.
- Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron*, 50, 507-517.
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45, 1363-1377.
- Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, 12, 366-375.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9, 357-381.
- Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology*, 92(2), 1144-1152.
- Balleine, B. W., & Dickinson, A. (1998). Goaldirected instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37, 407-419.
- Balleine, B. W., Killcross, A. S., & Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *Journal of Neuroscience*, 23(2), 666-675.
- Barense, M. D., Bussey, T. J., Lee, A. C., Rogers, T. T., Davies, R. R., Saksida, L. M., et al. (2005). Functional specialization in the human medial temporal lobe. *Journal of Neuroscience*, 25, 10239-10246.

- Bargh, J. A. (Ed.). (1990). Auto-motives: Preconscious determinants of social interaction. (Vol. 2). New York: Guilford Press.
- Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215-232). Cambridge, MA: MIT Press.
- Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron, 47, 129-141.
- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293-1295.
- Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology* (*Berl*), 191, 391-431.
- Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming*. Belmont, MA: Athena Scientific.
- Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through it role in working memory organization. *Journal of Neuroscience*, 26, 916-925.
- Bonardi, C., Graham, S., Hall, G., & Mitchell, C. (2005). Acquired distinctiveness and equivalence in human discrimination learning: Evidence for an attentional process. *Psychonomic Bulletin & Review*, 12, 88-92.
- Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, 281, 1185-1187.
- Buckmaster, C. A., Eichenbaum, H., Amaral, D. G., Suzuki, W. A., & Rapp, P. R. (2004). Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *Journal of Neuroscience*, 24, 9811-9825.

- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Science*, 11, 49-57.
- Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, 379, 255-257.
- Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience & Biobehavioral Reviews, 26, 321-352.
- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, 2, 844-847.
- Cohen, N. J. (1984). Preserved learning capacity in amnesia: Evidence for multiple memory systems. In L. Squire & N. Butters (Eds.), *The neuropsychology of memory* (pp. 83-103). New York: Guilford Press.
- Cohen, N. J., & Eichenbaum, H. (1993). Memory, amnesia, and the hippocampal system. Cambridge, MA: MIT Press.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207-210.
- Cohen, N. J., & Squire, L. R. (1981). Retrograde amnesia and remote memory impairment. Neuropsychologia, 19, 337-356.
- Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2003). L-dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with parkinson's disease. *Neuropsychologia*, 41, 1431-1441.
- Corbit, L. H., & Balleine, B. W. (2000). The role of the hippocampus in instrumental conditioning. *Journal of Neuroscience*, 20, 4233-4239.
- Corbit, L. H., & Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research*, 146, 145-157.
- Daselaar, S. M., Fleck, M. S., Prince, S. E., & Cabeza, R. (2006). The medial temporal lobe distinguishes old from new independently of consciousness. *Journal of Neuroscience*, 26, 5835-5839.
- Daw, N. D., Courville, A. C., & Touretzky, D. S. (2006). Representation and timing in

- theories of the dopamine system. *Neural Computation*, *18*, 1637-1677.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8, 1704-1711.
- Daw, N. D., Niv, Y., & Dayan, P. (2006). Actions, values, policies and the basal ganglia. In
 E. Bazard (Ed.), Recent breakthroughs in basal ganglia research (pp. 111-130). New York: Nova Science Publishers.
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441, 876-879.
- Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36, 285-298.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8, 1611-1618.
- Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, 84, 3072-3077.
- DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*, 13, 281-285.
- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge and New York: Cambridge University Press.
- Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. In C. R. Gallistel (Ed.), *Stevens' handbook of experimental psychology* (Vol. 3). New York: Wiley.
- Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Animal Learning and Behavior*, 23, 197-206.
- Dobbins, I. G., Rice, H. J., Wagner, A. D., & Schacter, D. L. (2003). Memory orientation and success: Separable neurocognitive components underlying episodic recognition. *Neuropsychologia*, 41, 318-333.
- Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the*

National Academy of Sciences (US), 94, 7109-7114.

- Eichenbaum, H., Stewart, C., & Morris, R. G. (1990). Hippocampal representation in place learning. *Journal of Neuroscience*, 10, 3531-3542.
- Eichenbaum, H. E., & Cohen, N. J. (2001). From conditioning to conscious recollection: Memory systems of the brain. New York: Oxford University Press.
- Eitam, B., Hassin, R. R., & Schul, Y. (2008). Non-conscious goal pursuit in novel environments: The case of implicit learning. *Psychological Science*, 19, 261-267.
- Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*, 1481-1489.
- Faure, A., Haberland, U., Condé, F., & El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *Journal of Neuroscience*, 25, 2771-2780.
- Ferbinteanu, J., Kennedy, P. J., & Shapiro, M. L. (2006). Episodic memory—from brain to mind. *Hippocampus*, 16, 691-703.
- Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299, 1898-1902.
- Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings* of the National Academy of Sciences (US), 103, 11778-11783.
- Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *Journal of Cognitive Neuroscience*, 17, 51-72.
- Frank, M. J., Seeberger, L. C., & O'Reilly R, C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, 306, 1940-1943.
- Fuster, J. M. (1997). The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe (3rd ed.). Philadelphia: Lippincott-Raven.
- Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. Annual Review of Psychology, 49, 87-115.
- Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus rep-

- resentation: A computational theory. *Hippocampus*, *3*, 491-516.
- Goldman-Rakic, P. S., Selemon, L. D., & Schwartz, M. L. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience*, 12, 719-743.
- Graybiel, A. M., Hirsch, E. C., & Agid, Y. (1990). The nigrostriatal system in Parkinson's disease. Advances in Neurology, 53, 17-29.
- Grice, G. R., & Davis, J. D. (1958). Mediated stimulus equivalence and distinctiveness in human conditioning. *Journal of Experimental Psychology*, 55, 565-571.
- Grice, G. R., & Davis, J. D. (1960). Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. *Journal of Experimental Psychology*, 59, 391-395.
- Griffiths, D., Dickinson, A., & Clayton, N. (1999). Episodic memory: What can animals remember about their past? *Trends in Cognitive Science*, *3*, 74-80.
- Haber, S. N. (2003). The primate basal ganglia: Parallel and integrative networks. *Journal of Chemical Neuroanatomy*, 26, 317-330.
- Hall, G., Ray, E., & Bonardi, C. (1993). Acquired equivalence between cues trained with a common antecedent. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 391-399.
- Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract statebased inference during decision making in humans. *Journal of Neuroscience*, 26, 8360-8367.
- Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *Journal of Neurophysiology*, 95, 948-959.
- Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., et al. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience*, 24, 1660-1665.
- Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hip-

- pocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences (US)*, 104, 1726-1731.
- Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96, 651-656.
- Horvitz, J. C. (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behavioural Brain Research*,137, 65-74.
- Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. 215-232). Cambridge, MA: MIT Press.
- Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: The role of reward-related learning and memory. *Annual Review of Neuroscience*, 29, 565-598.
- Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience*, 24, 7540-7548.
- Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: Implications for models of recognition memory. *Journal of Neuroscience*, 24, 4172-4180.
- Kakade, S., & Dayan, P. (2002). Dopamine: Generalization and bonuses. *Neural Networks*, 15, 549-559.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the American Statistical Association, 90, 773-795.
- Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. Cerebral Cortex, 13, 400-408.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308, 78-83.
- Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *Journal of Neuroscience*, 20, 6173-6180.

- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. Science, 273, 1399-1402.
- Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., et al. (2005). Perceptual deficits in amnesia: Challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia*, 43, 1-11.
- Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. Forgas, K. Williams, & W. Von Hippel (Eds.), Social judgments: Implicit and explicit processes (pp. 44-67). New York: Cambridge University Press.
- Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., et al. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, 125, 624-639.
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2), 339-346.
- McClure, S. M., Daw, N. D., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, 26, 423-428.
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306, 503-507.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*, 44, 379-387.
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16, 1936-1947.
- Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, 43, 133-143.
- Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience*, *9*, 1057-1063.
- Myers, C. E., & Gluck, M. A. (1996). Corticohippocampal representations in simul-

taneous odor discrimination: A computational interpretation of Eichenbaum, Mathews, and Cohen (1989). *Behavioral Neuroscience*, 110, 685-706.

- Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., et al. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience*, 15, 185-193.
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, 191, 507-520.
- Niv, Y., Duff, M. O., & Dayan, P. (2005). Dopamine, uncertainty and TD learning. Behavioral and Brain Functions, 1, 6.
- Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Science*, 10, 375-381.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304, 452-454.
- O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. Current Opinion in Neurobiology, 14, 769-776.
- O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, 38, 329-337.
- Otten, L. J., Henson, R. N., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain*, 124, 399-412.
- Owen, A. M. (1997). Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives. *Progress in Neurobiology*, 53, 431-450.
- Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience*, 9, 1465-1472.
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience

- into memory. *Trends in Cognitive Science*, *6*, 93-102.
- Parkinson, J. A., Dalley, J. W., Cardinal, R. N., Bamford, A., Fehnert, B., Lachenal, G., et al. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: Implications for mesoaccumbens dopamine function. Behavioural Brain Research, 137, 149-163.
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442, 1042-1045.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature*, 414, 546-550.
- Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, 41, 245-251.
- Port, R. L., & Patterson, M. M. (1984). Fimbrial lesions and sensory preconditioning. *Behavioral Neuroscience*, *98*, 584-589.
- Preston, A. R., & Gabrieli, J. D. (2008). Dissociation between explicit memory and configural memory in the human medial temporal lobe. *Cerebral Cortex*, 18, 2192-2207.
- Redgrave, P., & Gurney, K. (2006). The shortlatency dopamine signal: A role in discovering novel actions? *Nature Reviews Neuroscience*,7, 967-975.
- Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends* in *Neurosciences*, 22, 146-151.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton Century Crofts.
- Robbins, T. W. (1996). Refining the taxonomy of memory. *Science*, 273, 1353-1354.
- Robbins, T. W., & Everitt, B. J. (2007). A role for mesencephalic dopamine in activation: Commentary on Berridge (2006). *Psychopharmacology (Berl)*, 191, 433-437.

- Robinson, S., Sandstrom, S. M., Denenberg, V. H., & Palmiter, R. D. (2005). Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. *Behavioral Neuroscience*, 119, 5-15.
- Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, 10, 1615-1624.
- Ryan, J. D., & Cohen, N. J. (2004). Processing and short-term retention of relational information in amnesia. *Neuropsychologia*, 42, 497-511.
- Salamone, J. D. (2007). Functions of mesolimbic dopamine: Changing concepts and shifting paradigms. *Psychopharmacology* (Berl), 191, 389.
- Satoh, T., Nakai, S., Sato, T., & Kimura, M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of Neuroscience*, 23, 9913-9923.
- Schacter, D. L., & Addis, D. R. (2007a). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society of London* (B Biological Sciences), 362, 773-786.
- Schacter, D. L., & Addis, D. R. (2007b). Constructive memory: The ghosts of past and future. *Nature*, 445, 27.
- Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fmri and pet studies of episodic encoding and retrieval. *Hippocampus*, *9*, 7-24.
- Schnyer, D. M., Dobbins, I. G., Nicholls, L., Schacter, D. L., & Verfaellie, M. (2006). Rapid response learning in amnesia: Delineating associative learning components in repetition priming. *Neuropsychologia*, 44, 140-149.
- Schultz, W. (1998). The phasic reward signal of primate dopamine neurons. *Advances in Pharmacology*, 42, 686-690.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275, 1593-1599.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Shohamy, D., Myers, C. E., Geghman, K. D., Sage, J., & Gluck, M. A. (2006). L-dopa impairs learning, but spares gen-

- eralization, in Parkinson's disease. *Neuropsychologia*, 44, 774-784.
- Shohamy, D., Myers, C. E., Grossman, S., Sage, J., & Gluck, M. A. (2005). The role of dopamine in cognitive sequence learning: Evidence from Parkinson's disease. Behavioural Brain Research, 156, 191-199.
- Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feed-back-based learning: Converging data from neuroimaging and neuropsychology. *Brain*, 127, 851-859.
- Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilstic category learning. *Neuroscience and Biobehavioral Reviews*, 32, 219-236.
- Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience*, 118(4), 676-686.
- Shohamy, D., & Wagner, A. D. (2007). *Medial temporal lobe and basal ganglia contributions to learning and flexible transfer.* Paper presented at the Cognitive Neuroscience Society.
- Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*,4, 637-648.
- Squire, L. R. (1987). The organization and neural substrates of human memory. *International Journal of Neurology*, 21-22, 218-222.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195-231.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. Personality and Social Psychology Review, 8, 220-247.
- Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3, 9-44.
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. Cambridge, MA: MIT Press.
- Suzuki, W. A., & Amaral, D. G. (2004). Functional neuroanatomy of the medial temporal lobe memory system. *Cortex*, 40, 220-222.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences (US)*, 104, 642-647.

- Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits corticobasal ganglia loops. *Nature Neuroscience*, 7, 887-893.
- Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron*, 41, 281-292.
- Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *Journal of Neuroscience*, 27, 4019-4026.
- Wagner, A. D., & Davachi, L. (2001). Cognitive neuroscience: Forgetting of things past. *Current Biology*, 11, R964-R967.
- Wagner, A. D., Koutstaal, W., & Schacter, D. L. (1999). When encoding yields remembering: Insights from event-related neuroimaging. Philosophical Transactions of the Royal Society of London (B Biological Sciences), 354, 1307-1324.
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al.

- (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281, 1188-1191.
- Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality* and Social Psychology, 82, 919-934.
- White, N. M. (1997). Mnemonic functions of the basal ganglia. *Current Opinion in Neurobiology*, 7, 164-169.
- Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Duzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron*, 45, 459-467.
- Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. European Journal of Neuroscience, 19, 181-189.
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. European Journal of Neuroscience, 22, 513-523.