Learning and memory systems supporting
decision making in the human brain

G. Elliott Wimmer

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2012
ABSTRACT

Learning and memory systems supporting decision making in the human brain

G. Elliott Wimmer

We successfully navigate the world by making decisions based on what we have learned. In the brain, two prominent learning systems have been identified and each is likely to guide decisions in different ways. Research on decision making has primarily focused on a system for reward learning in the striatum. These studies have illuminated the how repeated choices and rewards build representations that guide choices and actions when encountering the same situation again. However, in a constantly changing environment, choices may not repeat themselves. Further, the environment may have more structure than simple reward learning can navigate. In these situations, decisions may be guided by a different learning system, namely a flexible learning system in the hippocampus which encodes episodes, or more broadly, relations between stimuli. However, investigations into the role of a reward learning system and a relational learning system in decision making have developed largely independently of each other. In the studies described below, I explore the function and interactions of these learning systems in value-guided decision making. Complementarily, I also explore how ongoing
reward learning may modulate memory formation in the hippocampal system. In these studies, I demonstrate that reward learning and decision making is influenced by relational learning, and that these effects are predicted by hippocampal-striatal connectivity during learning. Separately, I establish that episodic memory is, in turn, influenced by ongoing reward learning. Successful memory is predicted by modulations of reward and memory regions including the striatum and hippocampus. Overall, these results provide novel insights into the learning systems encoding memories for successful decision making in the future.
# Table of Contents

Table of Contents

List of Charts, Graphs, and Illustrations

Acknowledgements

Chapter 1: Introduction

I. Introduction

II. Systems for learning

III. Interactions between learning systems

IV. Systems for decision making

V. Questions and chapter introductions

Chapter 2: Generalization of value in reinforcement learning by humans

Introduction

Materials and Methods

Results

Discussion

Chapter 3: Preference by Association: A Neural Mechanism for How Memory Biases Decisions

Main Text

Materials and Methods
Supplementary Text

Chapter 4: Episodic memory modulation by reward learning

Introduction

Methods

Results

Discussion

Chapter 5: Choice and reward modulate episodic memory in the human brain

Introduction

Methods

Results

Discussion

Chapter 6: General discussion

Discussion

Limitations and Future Directions

References
List of Charts, Graphs, and Illustrations

Figure 2.1. Reward equivalence task .............................................................. 30
Table 2.1. Reinforcement learning model fits. ..................................................... 48
Figure 2.2. Ventral striatum BOLD and value generalization .............................. 53
Figure 2.3. Hippocampal BOLD correlated with value ....................................... 55
Figure 2.4. Ventral striatum-hippocampus connectivity related to generalization .... 57
Figure 3.1. Decision bias task ............................................................................ 71
Figure 3.2. Decision bias behavior and bias-predicting hippocampal activity .......... 74
Figure 3.3. Visual reactivation predicts bias ........................................................ 76
Figure 3.4. Hippocampus-caudate connectivity predicts bias ............................... 78
Figure S3.1. Liking ratings and association memory ........................................... 109
Figure S3.2. Association phase activation in the hippocampus ......................... 110
Figure S3.3. Reward learning-related activation ................................................ 111
Figure S3.4. Association phase group activation to visual stimuli ....................... 112
Figure S3.5. Association phase individual activation to visual stimuli ............... 113
Figure 4.1. Reward learning and memory task .................................................. 121
Table 4.1. Reward learning behavior ................................................................. 131
Figure 4.2. Reward influence on memory ........................................................... 132
Figure 4.3. Choice influence on memory ............................................................. 134
Table 4.2. Study 1: Choice stimulus memory prediction ........................................ 136
Table 4.3. Study 2: Feedback stimulus memory prediction ..................................... 137
Figure 5.1. Reward learning and memory task .................................................... 150
Figure 5.2. Memory prediction .......................................................................... 162
Figure 5.3. Activation predicting memory for chosen options. .......................... 163

Figure 5.4. Memory and prediction error interaction in the right putamen. .......... 165

Figure 5.5. Memory and prediction error interaction in the right caudate................... 166

Figure 5.6. Reward memory effect correlates with caudate interaction of prediction error and memory. ......................................................... 168

Figure S5.1. Hippocampus and memory......................................................... 175

Figure S5.2. Occipital and parietal cortex and memory. ................................ 175

Figure S5.3. Choice and feedback activation............................................... 176
Acknowledgements

To my advisor, Daphna Shohamy, for invaluable mentorship and guidance over the course of my years at Columbia. The advice given when she put on her “advisor hat” has always ended up working out.

And to Nathaniel Daw, for giving me all kinds of modeling knowledge. A special thanks to members of the Daw lab for reliable stimulation.

To Brian Knutson, my pre-Ph.D. mentor. Without his infectious enthusiasm and support, I wouldn’t have succeeded in becoming a neuroscientist.

To the department at Columbia. Professors Elke Weber, Kevin Ochsner, Ed Smith, and Tor Wager. Especially the Learning Lab: the lab started when I arrived, and in the meantime, it’s grown into something pretty wonderful. I’d like to especially thank Erin Kendall Braun for putting up with random outbursts when new results came in and my co-Ph.D. students in the lab, Jenna Reinen and Juliet Davidow, for providing kind words and ideas along the way.

To friends from Berkeley now on the East Coast, especially Yao Wu, for support, food, drink, and good times.

Finally, to my family, who keep on supporting me despite the fact their neither they nor I know what I’m actually doing in my day job. My brother has unknowingly provided key
support by also becoming a science nerd here on the East Coast – although his science leads to making real things. My parents have provided endless knowledge and enthusiasm throughout my education. I wouldn’t have made it here without them.
Chapter 1: Introduction
I. Introduction

How do we make decisions? Making good decisions, from what to eat for lunch to what career to pursue, is essential for survival and success. A full understanding of how the brain learns from experience and makes future decisions is a pressing issue in psychology and neuroscience, as understanding these capacities can help us understand everyday actions, and most importantly, make progress in treating dysfunctions of behavior caused by addiction, disease, and psychiatric disorders. Research in neuroscience has started to unravel how the brain learns from experience and uses this knowledge to make choices. While this research is just beginning, we now have clues about the kinds of learning and memory that support decision making. In the research described in this dissertation, I illustrate several directions for expanding our understanding of learning and decision making by exploring how two different learning and memory systems, a reward learning system in the striatum and a relational memory system in the hippocampus, can guide our choices.

As we navigate the world, we are constantly encoding relationships between events, while at the same time, we are learning from occasional experiences that involve reward or punishment. Navigating the world establishes flexible relational representations of the environment via learning processes in the hippocampus and related structures (for review, see Eichenbaum and Cohen, 2001). Meanwhile, experiences of reward or punishment are encoded as more simple stimulus-reward associations by a learning system in striatum and other brain regions (for review, see Rangel et al., 2008). While these systems are traditionally studied separately, a representation of the world in
conjunction with knowledge of reward associations in can be exploited for later planning and goal-directed behavior to acquire rewards and avoid punishments (Tolman, 1948).

Decision making in everyday life complex, and to deal with the complexity of our environment, we utilize multiple types of information to make decisions. Studies of learning from reward feedback have thus far made the most progress in detailing a source of information for decision making. This research has shed light on the important role of the striatum and dopamine in learning and choice, but many open questions and unexplored areas remain. In everyday experience, instead of well-learned reward associations, we often use more flexible memory for experiences, especially in novel situations. This suggests a prominent role for the hippocampus in decision making. However, the role of relational memory and the hippocampus in decision making has so far received little attention. Further, reward learning and relational learning processes and the supporting neural systems are often studied in relative isolation (e.g., Davachi, 2006; Schultz, 2006). Intriguingly, recent evidence suggests that relational memory in the hippocampus interacts with reward learning in the striatum to allow generalization of information to novel situations (Shohamy and Adcock, 2010). A complete understanding of human learning and decision making may thus require a consideration of the joint and interactive contributions of these learning systems.

The research in this dissertation explores connections between the cognitive and neural systems underlying reward and relational learning in decision making. In the four studies described in this dissertation, I probe the interaction of reward learning and relational learning to understand how different types of learning experiences can influence later behavior. This approach attempts to connect the study of learning and
decision making more closely to the learning systems utilized in the dynamic environments encountered in everyday behavior, where there is no separation between reward or relational learning and where different learning and memory systems may be engaged at the same time. Complementarily, I investigate the impact of concurrent reward learning on episodic memory formation. First, in this chapter, I outline the background and current state of neuroscience research on the neural systems for reward and relational learning in decision making.

II. Systems for learning

A primary goal of research in neuroscience is to understand the neural systems that support learning and memory for future adaptive behavior. We have learned that the brain contains specialized systems for making decisions based on different types of information. I will focus on two systems that in particular have been shown to be prominent in encoding behaviorally-relevant information: a system for learning reward associations from repeated experience and a system for learning relations and episodic memories from a single experience. In the brain, a reward (or habit) learning system is associated with the basal ganglia, including the striatum and midbrain dopamine nuclei, while a relational (or episodic) system is associated with the hippocampus and surrounding medial temporal lobe (MTL) cortices (Squire, 1987; Schacter, 1990; Squire, 1992; Cohen and Eichenbaum, 1993; Knowlton et al., 1996; Gabrieli, 1998; Wagner et al., 1998; Eichenbaum and Cohen, 2001; Poldrack et al., 2001; Paller and Wagner, 2002; Hartley and Burgess, 2005; Foerde et al., 2006; Eichenbaum et al., 2007; Daw and Shohamy, 2008; Foerde and Shohamy, 2011b). The discovery of separate functions of a
reward learning system and a relational learning system several decades ago has supported the development of the prominent “multiple memory systems” perspective for understanding how behavior is driven by underlying neural systems.

While this overview of reward and relational learning and memory systems is simple, the cognitive and neural systems supporting these types of learning are complex. Behaviorally, decades of learning research in animals has revealed a multiplicity of stimulus-reward and stimulus-response representations with distinct influences on behavior (Daw et al., 2005; Dayan et al., 2006; Balleine et al., 2008; Rangel et al., 2008). The neural systems supporting learning from reward are also complex, and the regions involved undoubtedly extend beyond dopamine neurons and the striatum (Daw and Doya, 2006; Balleine et al., 2008; Rangel et al., 2008; Takahashi et al., 2009; Li et al., 2011; McDannald et al., 2011). For the purposes of the studies described in this dissertation, however, I will use “reward learning” as a general term for the representation and updating of stimulus-response or stimulus-reward (S-R) associations in the brain, where these types of associations are believed to depend on the striatum and its dopaminergic inputs.

Similarly, research on relational memory has revealed multiple kinds of representations (Eichenbaum and Cohen, 2001; Davachi, 2006; Eichenbaum et al., 2007). Neurally, in contrast to the multiple systems supporting reward learning, there is more convergence on the importance of a single structure, the hippocampus. The hippocampus, because of a combination of unique circuitry and high plasticity (McClelland et al., 1995), is consistently identified as a critical locus of memory formation. In humans, damage to the hippocampus and surrounding medial temporal lobe results in a highly
specific memory deficit that renders new episodic learning impaired while sparing other learning processes (Cohen and Squire, 1980, 1981; Squire, 1992; Gabrieli, 1998; Eichenbaum and Cohen, 2001). Functional magnetic resonance imaging (fMRI) studies have shown that hippocampal activity supports the successful encoding of experience (Brewer et al., 1998; Wagner et al., 1998; Schacter and Wagner, 1999; Kirchhoff et al., 2000; Otten et al., 2001). More specifically, activity predicts successful encoding of relations between items and between items and contexts (Eichenbaum and Cohen, 2001; Paller and Wagner, 2002; Davachi, 2006; Eichenbaum et al., 2007; Hannula and Ranganath, 2008). Activity in single neurons in the hippocampus has been shown to encode and recollect sequences of places in animals as well as humans (Foster and Wilson, 2006; Gelbard-Sagiv et al., 2008; Carr and Frank, 2012), supporting the encoding episodes.

In general, these results suggest that the memory representations in the hippocampus are relational, in that they contain information about spatial, temporal, or associative relations between multiple stimuli (Eichenbaum and Cohen, 2001; Eichenbaum et al., 2007; Staresina and Davachi, 2009). A second key feature of these memories is representational flexibility – hippocampus-based memories can be accessed, transferred, and used in novel contexts (Cohen, 1984; Cohen and Eichenbaum, 1993; Eichenbaum, 2000). Based on the critical role of the hippocampus in encoding relations, or stimulus-stimulus (S-S) associations, I will use “relational learning” as a general term for the representation this kind of learning.

Reward learning
Neuroscience research on decision making has focused extensively on the role of the learned reward associations. The learning of predictive relationships between stimuli and reward is well described by reinforcement learning theories (Rescorla and Wagner, 1972; Houk et al., 1995; Schultz et al., 1997; Frank et al., 2004; Everitt and Robbins, 2005; Daw and Doya, 2006; Schultz, 2006; Rangel et al., 2008). Behavioral and economic studies have demonstrated that choices are predominantly driven by learned value. Basing choices on value is obviously an adaptive strategy; it allows organisms to make choices that provide them with the food, money and other resources needed for survival.

In reward learning experiments, reinforcement learning models accurately predict the phasic spiking of dopamine neurons as well as the blood oxygen-level dependent (BOLD) signal from fMRI studies in the ventral striatum (Houk et al., 1995; Montague et al., 1996; Schultz et al., 1997; Berns et al., 2001; Knutson et al., 2001; McClure et al., 2003; O'Doherty et al., 2003). Dopamine neuron phasic spiking is well-described by the difference between the expected and received reward, the reward prediction error signal in reinforcement learning models. This dopamine response can then update the value of stimulus-reward associations that underlies value predictions and behavior (e.g. Reynolds et al., 2001; Tsai et al., 2009). The link between the key learning signal in reinforcement learning models and a neural correlate in a widely-projecting neuromodulatory signal has been one of the biggest success stories in systems and computational neuroscience.

While research in humans and animals on value learning has been extremely productive and influential, the basic reinforcement learning theories that this research relies upon are relatively limited in explanatory scope. Beyond stimulus-reward
associations, basic reinforcement learning theories do not explain how other forms of learning guide decision making (Tolman, 1948; Doya, 1999; Daw et al., 2005; Daw and Shohamy, 2008). While in many situations simple reward learning is sufficient, we have additional cognitive and learning systems to guide choice.

Reinforcement learning relies on repeated experiences to extract associations between predictive stimuli and rewarding or aversive feedback. This kind of learning is “model-free”, in that simple value associations are acquired without connections to other stimuli or a representation of the world (Daw et al., 2005). Learning over repeated experience is a powerful feature, because it allows an organism to easily navigate regularly experienced events by automatically extracting predictive relationships between stimuli and outcomes. Similarly, habitual motor behaviors offer a large advantage over re-planning every regularly used action. However, if the environment shifts or the animal’s internal motivation changes, reinforcement learning requires re-exposure to contingencies to re-learn the values of decision options.

The inflexibility of a simple reward learning system can be demonstrated by behavioral paradigms such as devaluation. For example, when a lever is reliably and repeatedly associated with a favorite food, pressing the lever to get a reward can become a habitual behavior. If the food reward were suddenly paired with illness, it would be adaptive for the animal to avoid the lever and explore other options. However, a reward learning system based on simple reinforcement learning will only update the value of the lever after re-experiencing the lever-food-illness association. This is not an adaptive behavior, and indeed, this is not what animals do. When the environment changes, animals have been shown to rely on a more flexible learning system that has access to a
representation of the environment, including predictive (S-S) associations between neutral stimuli (Daw et al., 2005). This different type of learning has been shown to depend on different non-habitual circuits including the dorsomedial striatum, the frontal cortex, and possibly the hippocampus (Yin and Knowlton, 2006). Lesions to components of this alternate learning system render behavior inflexible, while lesions to the lateral striatum prevent behavior from becoming habitual (for review, see Yin and Knowlton, 2006). Thus, learning and decision making in complex and shifting situations may in part rely on one or more separate learning systems distinct from the reinforcement learning-related striatal system.

Relational learning

One candidate learning system that could complement a reward learning system in decision making is a relational learning system localized to the hippocampus and surrounding MTL. Studies on episodic learning and memory have highlighted a central role for the hippocampus in flexible, relational learning (Knowlton et al., 1996; Dusek and Eichenbaum, 1997; Eichenbaum and Cohen, 2001; Davachi, 2006). Extensive converging evidence indicates that the hippocampus and surrounding MTL cortices support the rapid formation of memories for single episodes (Squire, 1987, 1992; Cohen and Eichenbaum, 1993; Gabrieli, 1998; Wagner et al., 1998; Eichenbaum and Cohen, 2001; Paller and Wagner, 2002; Eichenbaum et al., 2007).

While initial research in humans suggested that the hippocampus was critical for forming declarative memories characterized by conscious experience, more recent research suggests a role for the hippocampus in relational learning regardless of
awareness (Henke, 2010; Hannula and Greene, 2012). Research is converging on the view that the hippocampus plays an essential role in binding experiences together, whether parts of a scene or elements sequentially experienced in time (Davachi, 2006; Staresina and Davachi, 2009; Turk-Browne et al., 2010). This perspective on relational memory as separate from a question of awareness also facilitates an easier integration with research on hippocampal function in animals. Animal research, in contrast to most human memory research, has focused on the hippocampus as a critical region for representing for spatial location, navigation, and binding of elements in the environment (Eichenbaum and Cohen, 2001).

Until recently, little attention was given to a role for relational memories in the hippocampus in human decision making. This may have been partly because of a bias to view human hippocampal function from a declarative memory point of view. However, research on human relational learning in the hippocampus strongly suggests that it is a critical system for adaptive behavior, as will be discussed below. Stimulus-stimulus relational associations between stimuli can provide building blocks for planning and evaluating consequences in decision making (Lengyel and Dayan, 2005). A relational memory system in the hippocampus may thus be well suited to address the inflexibility of decision making based on reinforcement learning in the striatum (Daw and Shohamy, 2008).

III. Interactions between learning systems

A reward learning system in the striatum and a relational learning system in the hippocampus have been proposed to compete for the control of behavior (Poldrack and
Packard, 2003). Frequently, studies of memory system function are designed to isolate the roles of different learning systems. One commonly used paradigm is the Morris water maze, where animals can use either a location-based strategy (reliant on the hippocampus) or a cue-based strategy (reliant on the striatum) to solve the task. It has been shown that inactivating one learning system can improve the performance of the type of learning supported by the other learning system (e.g. Lee et al., 2008). In general, studies combining lesions of either the MTL or the striatum with behavioral tasks that probe either spatial-relational learning or habit learning have shown that an intact MTL is essential for the former, while the striatum (specifically, the caudate) is essential for the latter (Packard et al., 1989; Kesner et al., 1993; McDonald and White, 1993; Packard and McGaugh, 1996; Packard, 1999; for a review, see Poldrack and Packard, 2003).

In humans, striatal disruption (such as occurs in Parkinson’s disease) impairs performance on a variety of incremental, stimulus-response learning tasks (Saint-Cyr et al., 1988; Downes et al., 1989; Owen et al., 1993; Knowlton et al., 1996; Swainson et al., 2000; Shohamy et al., 2004a; Shohamy et al., 2004b; Shohamy et al., 2005; Shohamy et al., 2006), but spares performance on tasks that involve relational memory such as declarative knowledge (Knowlton et al., 1996). In a related human functional imaging study, a negative correlation was found between hippocampal and striatal activation during learning (Poldrack et al., 2001); however, with fMRI data, a causal connection between interactions and behavior cannot be established.

These studies and many others suggest that reward and relational learning not only support different kinds of learning but also that in some situations these systems competitively interact (Poldrack and Packard, 2003).
Interaction via hippocampal-striatal connectivity

A competitive interaction between reward learning and relational learning may not characterize all interactions, however, and many studies now support a case for some functional cooperation between these systems (for review, see Shohamy and Adcock, 2010; Pennartz et al., 2011). Neurophysiological studies in animals have shown that activity in the hippocampus and striatum is correlated during learning as well as in offline replay (Pennartz et al., 2004; Johnson and Redish, 2007; Lansink et al., 2008; Tort et al., 2008; Lansink et al., 2009; Singer and Frank, 2009; van der Meer and Redish, 2009).

Cooperative interactions between systems could be supported by anatomical connections between the hippocampus and ventral striatum (Kelley and Domesick, 1982; Cohen et al., 2009). Neuroanatomical studies have shown that the hippocampus projects directly to the ventral striatum, forming a component of the “reward-related” ventral corticostriatal loop (Haber et al., 2006; Cohen et al., 2009; Haber and Knutson, 2010). Neurophysiological studies suggest that this connection underlies functional interactions between these two regions (Lansink et al., 2009; van der Meer and Redish, 2009).

Directly supporting a role for such connectivity in adaptive behavior, hippocampal inputs to the striatum have been shown to be critical for learning place preferences, a classic measure of reward learning (Ito et al., 2008). Further, the hippocampus has been shown to exert some control over striatal dopamine levels and striatal activity. For example, stimulation of the ventral hippocampus has been shown to enhance the number of spontaneously activated dopamine neurons in the midbrain,
resulting in significantly greater dopamine release in the ventral striatum (Legault and Wise, 1999; Lodge and Grace, 2006, 2008).

From a computational perspective, it may be beneficial for these systems to cooperate by sharing information. For example, if a relational learning system in the hippocampus is engaged in planning a course of action in an environment, it could be efficient for this system to access model-free values stored in the reward learning system. This type of interaction has been hinted at in several neurophysiological studies of maze navigation (Johnson and Redish, 2007; van der Meer and Redish, 2009). While rats navigated a maze, hippocampal activity at decision points reflected future paths in toward reward (Johnson and Redish, 2007). In the same task, activity in the striatum was increased at decision points, supporting the hypothesis that planning-like activity in the hippocampus accesses reward representations in the striatum (van der Meer and Redish, 2009). However, even if these systems share information, it is possible that they compete for the eventual control of behavioral output (e.g. Daw et al., 2005).

Common dopaminergic inputs to both systems

Cooperative interactions between reward and relational learning systems are also supported by the fact that both the striatum and the hippocampus receive neuromodulatory inputs from midbrain dopamine neurons. In addition to the well-known and prominent midbrain dopamine neuron projections to the striatum, dopamine neurons in the midbrain ventral tegmental area (VTA) also project to the hippocampus (Fuxe, 1965; Lindvall and Bjorklund, 1974; Swanson, 1982; Gasbarri et al., 1994a; Gasbarri et al., 1994b). (However, see Smith and Greene (2012) for data suggesting that the}
noradrenergic neurons from the locus coeruleus are the actual source of dopamine in the hippocampus. Dopamine has been shown to modulate long-term plasticity in the hippocampus. For example, dopamine agonists promote longer-term facilitation of synaptic transmission (Huang and Kandel, 1995; Otmakhova and Lisman, 1996; Lisman et al., 2011), while dopamine antagonists prevent it (Frey et al., 1990). Functionally, the hippocampus may also communicate with the dopamine system via a circuit connecting the hippocampus, ventral striatum, and VTA (Lisman and Grace, 2005).

Many studies support the behavioral importance of dopamine in the hippocampus. Lesions to midbrain dopamine projections to the hippocampus have been shown to impair spatial memory (Gasbarri et al., 1996). Dopamine in the hippocampus has been shown to be necessary for long-lasting maintenance of fear memories (Rossato et al., 2009). Increases in hippocampal plasticity in response to a novel environment depends on the activation of dopamine receptors (Li et al., 2003; Lemon and Manahan-Vaughan, 2006; Granado et al., 2008). Also, dopamine D1-type receptor knockout also alters the ability of hippocampal place cells to adapt coding to a new environment (Tran et al., 2008). Finally, Morris and colleagues have demonstrated that dopamine in the hippocampus is critical for the formation of lasting episodic-like memories of spatial organization of rewards (Bethus et al., 2010). Collectively, these studies support a tight linkage between the midbrain dopamine system and the hippocampus in memory formation. Thus, dopamine, which has been primarily associated with stimulus-response learning and value-based decision making, may play an essential role in hippocampal dependent episodic memory.
Overlapping learning signals in functional MRI studies

While fMRI studies in humans initially supported a competitive view of hippocampal and striatal function during learning (Poldrack et al., 2001), recent work has shown support for similar responses across systems. In the realm of reward learning and value-based decision making, multiple recent studies have found hippocampal activity correlated with value and feedback (Kumaran et al., 2009; Lebreton et al., 2009; Dickerson et al., 2011; Foerde and Shohamy, 2011a; Simon and Daw, 2011; Wimmer et al., 2012). In some studies, hippocampal activation has also been found to correlate with reward prediction errors during reward receipt, a result characteristically localized to the striatum (Dickerson et al., 2011; Foerde and Shohamy, 2011a). In other studies, hippocampal activation correlated with the value of options during choice, a result characteristically localized to the VMPFC (Simon and Daw, 2011; Wimmer et al., 2012). These results further challenge a simple competitive account of striatal and hippocampal memory systems.

In relational memory studies, similar co-activation of striatal and hippocampal systems has been reported. In a motivated memory-encoding paradigm, successful memory formation for high reward trials was predicted by increased activity in and correlation between the hippocampus and VTA (Adcock et al., 2006). In an incidental encoding study, reward value cues activated the striatum, and midbrain and hippocampal activity predicted successful memory formation (Wittmann et al., 2005). Outside of a reward context, striatal signals have also been shown to reflect successful memory encoding (Sadeh et al., 2010).
IV. Systems for decision making

How does reward learning and relational learning support decision making? Theoretical and experimental studies of reward learning are closely linked to decision making. The function of phasic dopamine neuron activity was understood via prior research in engineering and computer science on how to program machines to learn from feedback (Houk et al., 1995; Montague et al., 1996; Sutton and Barto, 1998). In contrast, decision making is rarely approached from the perspective of relational learning. While research on reward learning benefitted from an immediate connection to behavior, the connection between relational learning and decision making is much less clear. Nevertheless, it is clear that the representations acquired by these systems support adaptive behavior. Relational memory and reward learning may converge in decision making, as discussed below.

Reward learning and decision making

To understand human decision making, animal conditioning studies and basic behavioral economic choice studies served as a starting point (e.g. O'Doherty et al., 2003). It is now well established that in functional MRI studies, the ventral striatum responds to reward prediction errors (Hare et al., 2008). Other regions have been shown to respond to choice value during decision making, such as the ventromedial prefrontal cortex (VMPFC) (Daw et al., 2006; Hare et al., 2008; Chib et al., 2009). Based on many studies, well-established findings such as reward prediction error and value responses in the human brain can be leveraged to study decision making in more complex and changing environments. Interestingly, several recent studies of reward learning illustrate
how reward learning and decision making are impacted by relational, or “higher-order” information.

In one example, Daw et al. (2011) demonstrated that the striatal reward prediction error signal in humans is sensitive to relationships in the environment that simple reinforcement learning would be blind to. In a learning task, participants made sequential decisions at two stages. By using a multi-stage design, reward feedback responses in the brain can be probed for the influence of different learning systems: one which tracks simple stimulus-reward associations (in a model-free manner, as in simple reward learning) and one that is aware of the overall two-stage structure of the game (in a model-based manner, perhaps related to relational learning). If the reward learning system in the striatum only has access to knowledge derived from simple reward learning, reward prediction errors should only reflect the model-free feedback component. However, participant’s choice behavior and the striatal reward response was actually best described by a combination of simple stimulus-reward learning and model-based learning (Daw et al., 2011). This result demonstrates that one of the most basic neural correlates of reward learning is also influenced by additional structure in the environment.

In a second example, a recent electrophysiology study demonstrated that dopamine neurons respond to higher-order contingencies in the environment (Bromberg-Martin et al., 2010a). In a task where stimulus-response-reward associations switched repeatedly, Bromberg-Martin found that after unexpected reward omission, on the next choice, the dopaminergic reward prediction error signal had already systematically shifted to correctly predict reward feedback on the alternate stimulus. Remarkably, this
signal was reflected not only in dopamine neurons, but also in a circuit upstream of the midbrain including the habenula and globus pallidus (Bromberg-Martin et al., 2010a).

These studies illustrate that simple reward learning signals may reflect additional higher-information about relationships in the environment. However, it is not known what neural systems support the acquisition of this type of information. It is possible that a learning system in the striatum itself tracks this information, but it is more likely that this information comes from other sources. Often, these kind of flexible representations and are related to the prefrontal cortex (Miller and Cohen, 2001; Hampton et al., 2006). However, given the role of the hippocampus in relational learning the hippocampus may also play a critical role in supplying relational information to a reward learning system (Lengyel and Dayan, 2005).

*Relational learning and decision making*

Research on episodic and relational memory has approached the question of value and decision making from a different direction. Human memory studies have explored the neural systems supporting memory for positive and negative stimuli (Phelps, 2004; Murty et al., 2011). For decision making, it is almost of no question that people will choose to avoid negative stimuli and approach positive stimuli, should the situation warrant it, regardless of whether or not emotional stimuli are remembered. However, memories for positive and negative events can be very useful for decision making when memories can be linked into relational representations that allow people to predict the consequences of future actions. Of course, memory formation in the hippocampus is characterized by its relational nature (Davachi, 2006).
While simple relational encoding is often not considered from a decision making perspective, it is precisely these kind of relations that can be used to construct a model of the world (Lengyel and Dayan, 2005). Such representations can then be used in planning and decision making to predict the consequences of actions as in model-based theories of decision making (Daw et al., 2005; Lengyel and Dayan, 2005). Additionally, the role of the hippocampus in navigation, as established by decades of research in animals (Eichenbaum and Cohen, 2001), strongly suggests that the hippocampus may be critical for model-based learning.

Notably, the multiple studies that support a role for the hippocampus in relational learning can also be viewed as providing support for a role of the hippocampus in model-based learning and decision making. For example, one of the most commonly cited studies for the role of the hippocampus in relational memory uses a transitive inference paradigm (Dusek and Eichenbaum, 1997). In this paradigm, animals learn via repeated stimulus-reward associations to prefer A > B, C > D, and D > E. Following learning is a test with novel pairs, where transfer of knowledge leads to a preference of B > D. Hippocampal lesions impair performance on these novel pairs, but not on pre-trained pairs (Dusek and Eichenbaum, 1997). Importantly, hippocampal damage specifically impairs the ability to perform such transfer, without significantly impacting the ability to learn the individual associations (Bunsey and Eichenbaum, 1996; Dusek and Eichenbaum, 1997; Eichenbaum and Cohen, 2001; Buckmaster et al., 2004). Interestingly, damage to the striatum results in the opposite pattern – impaired feedback-based learning of the individual associations, but spared transfer (Myers et al., 2003; Shohamy et al., 2006).
Another key set of studies for establishing a role of the hippocampus in relational memory also involves reward learning, and highlights a possible mechanism by which the hippocampus can influence decision making. In this paradigm, called acquired equivalence, participants learn a series of associations using trial and error. Each of the associations in the series is learned individually; however, the associations sometimes overlap between stimuli. For example, participants learn that A→X, B→X, and A→Y. Participants are then tested on transfer of this knowledge to a novel association (B→Y).

Animal studies suggest an important role for the hippocampus in transfer (Coutureau et al., 2002). In human participants with amnesic mild cognitive impairment, a condition that often leads to Alzheimer’s disease, the transfer across associations is impaired, but basic learning of associations through reward is intact (Myers et al., 2003). A study using functional MRI has shown that such transfer is supported by hippocampal activation, in conjunction with activation in the dopaminergic midbrain (Shohamy and Wagner, 2008). These results indicated that during learning, when individual associations were being experienced, the hippocampus was engaged in integrative encoding (Shohamy and Wagner, 2008).

These findings provide insight into a possible mechanism by which relational learning can influence decision making. By demonstrating that hippocampal activity during learning predicts subsequent choice, these findings raise the possibility that associative mechanisms in the hippocampus continuously integrate episodes as they are being experienced. This suggests that hippocampal-dependent decision making is essentially a form of generalization, emerging from relational links between learned representations driven by the overlap between them.
As described above, prominent support for the hippocampus in flexible relational learning comes from paradigms that involve coincident learning of reward associations. Such concurrent learning may more closely relate to everyday experience, where reward learning and relational learning are not artificially separated. Such findings cut against a view that proposes competition as a basic model of interaction between relational and reward learning (Poldrack and Packard, 2003). By learning relational representations that can serve as a basis for generalizing related reward experiences, the brain may be able to take advantage of relational learning to flexibly apply knowledge in novel decisions.

In conclusion, the research outline above has suggested an important role for relational learning in decision making. Studies of reward learning are expanding beyond simple probes of reinforcement learning and have so far demonstrated that classic reward signals in the brain are sensitive to relations in the environment that go beyond simple stimulus-response associations (Bromberg-Martín et al., 2010a; Daw et al., 2011). At the same time, studies of relational learning have demonstrated a critical role of the hippocampus in reward learning and flexible decision making (Dusek and Eichenbaum, 1997; Shohamy and Wagner, 2008).

V. Questions and chapter introductions

In this dissertation, I present four studies. Collectively, these studies aim to increase our understanding of the cognitive and neural systems that support learning and decision making, with a focus on two specific systems: a relational learning system in the hippocampus and a reward learning system in the striatum. In Chapter 2, I investigate
whether relational memory processes impact reward learning (Wimmer, Daw, and Shohamy, 2012). Specifically, I explore whether value generalizes across related stimuli and use fMRI to explore the neural correlates of this effect. While simple aspects of stimulus-reward learning have been extensively investigated, the environment often presents more complex situations that are outside the scope of basic reinforcement learning. In this experiment, my goal was to investigate the following questions: How is knowledge about value similarity between stimuli acquired? How does such knowledge impact choice behavior? Do the striatum and the hippocampus each play independent or cooperative roles in contributing to these different aspects of learning?

In Chapter 3, I investigate whether relational memory processes in the hippocampus interact with subsequent reward learning to shift value and bias decision making (Wimmer and Shohamy, in press). In the study described in Chapter 2, learning relations between stimuli and between stimuli and feedback was intermixed, and thus it was not possible to isolate the precise mechanisms underlying the transfer of reward across stimuli. In Chapter 3, we extend these findings with a new study aimed to elucidate the specific mechanism by which the hippocampus contributes to value assessment and reward learning, and how the hippocampus interacts with the striatum to support this process. This study focuses on the following questions: How does reward learning interact with relational associations? Can value transfer from a rewarded stimulus to a previously associated neutral stimulus? Does a cooperative interaction between relational learning in the hippocampus and reward learning in the striatum during reward learning supports later decision making?
Together, Chapters 2 and 3 address questions about how memory representations in the hippocampus interact with reward learning to guide choice. In the studies described in Chapter 4 (Wimmer and Shohamy, *in preparation*) and Chapter 5 (Wimmer, Braun, and Shohamy, *in preparation*), I turn to explore how episodic memory is impacted by choice and reward learning. In Chapter 4, I determine whether episodic memory is modulated by choice and reward in a dynamic reward learning environment. I also probe the effect of consolidation on this influence. In Chapter 5, I use fMRI to explore the neural mechanisms underlying the effect of choice and reward on memory. Specifically, I ask whether value signals in the VMPFC are also related to memory, and whether memory and is related to the striatal reward prediction error learning signal.
Chapter 2: Generalization of value in reinforcement learning by humans

In Chapter 2, I approach the use of relational representations in reward learning and choice by probing learning in a basic reinforcement learning task with a relational structure that links the reward likelihood of paired options.

Abstract

Research in decision making has focused on the role of dopamine and its striatal targets in guiding choices via learned stimulus-reward or stimulus-response associations, behavior that is well-described by reinforcement learning (RL) theories. However, basic RL is relatively limited in scope and does not explain how learning about stimulus regularities or relations may guide decision making. A candidate mechanism for this type of learning comes from the domain of memory, which has highlighted a role for the hippocampus in learning of stimulus-stimulus relations, typically dissociated from the role of the striatum in stimulus-response learning. Here, we used fMRI and computational model-based analyses to examine the joint contributions of these mechanisms to RL. Humans performed an RL task with added relational structure, modeled after tasks used to isolate hippocampal contributions to memory. On each trial participants chose one of four options, but the reward probabilities for pairs of options were correlated across trials. This (uninstructed) relationship between pairs of options potentially enabled an observer to learn about options’ values based on experience with the other options and to generalize across them. We observed BOLD activity related to learning in the striatum and also in the hippocampus. By comparing a basic RL model to one augmented to allow feedback to generalize between correlated options, we tested whether choice behavior and BOLD activity were influenced by the opportunity to generalize across correlated
options. Although such generalization goes beyond standard computational accounts of RL and striatal BOLD, both choices and striatal BOLD were better explained by the augmented model. Consistent with the hypothesized role for the hippocampus in this generalization, functional connectivity between the ventral striatum and hippocampus was modulated, across participants, by the ability of the augmented model to capture participants’ choice. Our results thus point toward an interactive model in which striatal RL systems may employ relational representations typically associated with the hippocampus.
Introduction

Research in decision making posits a computational role for the dopamine system and its striatal targets in guiding choices via learned stimulus-reward or stimulus-response associations (Houk et al., 1995; Schultz et al., 1997; Frank et al., 2004; Everitt and Robbins, 2005; Daw and Doya, 2006; Schultz, 2006; Rangel et al., 2008). However, there has been increasing recognition that this narrow mechanism for “habit” learning cannot explain the full diversity of choice behavior, or even the striatum’s contribution to it (Balleine et al., 2008; Rangel et al., 2008; Redish et al., 2008). Still, it remains less precisely understood how other forms of learning, possibly involving distinct cognitive and neural systems, contribute to choice (Doya, 1999; Daw et al., 2005; Daw and Shohamy, 2008).

One promising avenue for addressing this gap is the largely separate domain of memory research, where a finely detailed distinction between different forms of learning has long been established (Schacter, 1990; Squire, 1992; Knowlton et al., 1996; Gabrieli, 1998; Eichenbaum and Cohen, 2001). Perhaps the best-characterized system is that for episodic memory, associated with the hippocampus and operationally distinguished from a striatal habit system (Schacter, 1990; Squire, 1992; Knowlton et al., 1996; Gabrieli, 1998; Eichenbaum and Cohen, 2001; Poldrack et al., 2001; Hartley and Burgess, 2005; Foerde et al., 2006; Mattfeld and Stark, 2010). Echoing non-habitual accounts of decision making, hippocampal memories represent the relation between multiple arbitrarily associated stimuli. Due to their relational nature, hippocampal memories are also flexible and can be generalized across stimuli and contexts (Cohen and Eichenbaum, 1993; Dusek
and Eichenbaum, 1997; Eichenbaum and Cohen, 2001; Davachi, 2006; Shohamy et al., 2008; Staresina and Davachi, 2009).

In memory tasks, the relational hallmark of hippocampal memories has been demonstrated using procedures that first embed relations among stimuli and then probe whether later choices reflect relational knowledge (Dusek and Eichenbaum, 1997; Myers et al., 2003; Preston et al., 2004; Greene et al., 2006; Shohamy et al., 2006; Shohamy and Wagner, 2008; Zeithamova and Preston, 2010). For example, in ‘acquired equivalence,’ people first learn that stimulus A is associated with outcome X, and that stimulus B is also associated with outcome X. Having indirectly learned that A and B are related, in terms of their common outcome X, people later transfer additional knowledge about stimulus A to stimulus B, presumably based on the learned ‘equivalence’ between them. Converging evidence suggests that acquired equivalence depends on the hippocampus and surrounding medial-temporal lobe cortex (e.g., Coutureau et al., 2002; Myers et al., 2003; Shohamy and Wagner, 2008).

Here, we sought to leverage this approach in the context of a reinforcement learning task to determine whether relational encoding contributes to decision making. Participants made repeated choices in a reward learning task, in which the probability of reward associated with each of four options diffused randomly. Structured relationships between options’ outcomes were embedded via correlated reward probabilities between pairs of options, creating an (uninstructed) equivalence between them (Figure 2.1). Thus, this task incorporates one of the essential elements of ‘acquired equivalence’ (Honey and Hall, 1989; Myers et al., 2003; Shohamy and Wagner, 2008), namely, that pairs of options are related in virtue of sharing a common outcome, enabling (if this structure is
detected and encoded) generalization of subsequent learning between them. However, in contrast to studies in the memory domain, the common outcome here is a correlated likelihood of reward, rather than a particular stimulus. Moreover, this correlational structure is embedded within a trial-and-error reward learning task, allowing us to ask whether and how inferred similarity relationships of this kind can affect instrumental choice behavior. Importantly, standard reinforcement models (ranging from Thorndike’s (1911) law of effect to more modern TD rules (e.g. Schultz et al., 1997)) should in principle be entirely blind to this kind of structure.

We characterized learning behavior using reinforcement learning models in order to measure the extent to which choices are driven by the correlational structure across option’s values. We then used fMRI to identify regions of the brain where activation covaried with decision variables from the models, to investigate whether the inclusion of this structure implicated the hippocampus instead of (or in addition to) traditional reinforcement learning activations in the striatum. Critically, we could then examine these signals to test whether they reflected value generalization, and specifically whether ventral striatal BOLD activity reflected relational knowledge. Finally, we used multivariate analyses of the fMRI data to examine whether the use of such structure to guide choices might be reflected in increased functional connectivity between the hippocampus and the striatum.
Figure 2.1. Reward equivalence task.

Design of the reward equivalence paradigm. A, On each trial, participants chose one of four face options. After a delay, the outcome ($0.25 or $0.00) was revealed. In colored brackets, one example of option pairing is indicated. B, Drifting reward probability distribution defining the reward equivalence for one example pairing (left). Trial-by-trial reinforcement learning variables for 50 trials from an example participant: fMRI model regressors for prediction error (black) and prediction error difference due to generalization (red), and an illustration of a full generalization model prediction error (grey) (right).
Materials and Methods

Participants. Twenty-four right-handed fluent English speakers with normal or corrected-to-normal vision participated in the study. All participants were free of neurological or psychiatric disorders and fully consented to participate. Informed consent was obtained in a manner approved by the New York University Committee on Activities Involving Human Subjects. Three participants’ data were excluded: two due to software problems (one for a partial loss of behavioral responses, one for missing timing information), and a third because the participant elected to leave the experiment before the completion of data acquisition. Behavioral and functional imaging data are presented from the remaining twenty-one participants (mean age, 19.3 years; range, 18-28; 10 female). Participants were paid $20 per hour for the approximately 2-hour duration of participation plus one-fifth of the nominal rewards the participant earned in the experimental task.

Task. In the experimental task (Figure 1a; Daw and Shohamy, 2008), on each of 300 trials, participants chose one of four presented face stimuli and then received monetary feedback. This reinforcement learning task is a variant of a “four-armed bandit” task (Daw et al., 2006; Wittmann et al., 2008). The face stimuli, which were constant across trials and participants, were taken from the Stanford Face Database. The location of the faces was permuted randomly from trial to trial.

On each trial, participants had 2 s to choose between the four options (Figure 2.1A), using an MR-compatible button response pad held in the right hand. After the participant made a selection and until the end of the choice period, the selected option
was framed in blue and the unchosen options were decreased in brightness. Participants then received binary reward feedback for 2 s, a $0.25 “win” outcome represented by an image of a quarter dollar and a $0.00 “miss” outcome represented by a phase-scrambled image of a quarter dollar (Figure 2.1A). If no choice was recorded during the choice period, no reward outcome was displayed and the face options remained on the screen until the end of the trial. Trials were intermixed with variable duration inter-trial fixation null events (ITI; mean 2 s, range 0-12 s). The total time allotted for null events was equal to one-third of the scan time. The duration and distribution of null events was optimized for estimation of rapid event-related fMRI responses as calculated using Optseq software (http://surfer.nmr.mgh.harvard.edu/optseq/). The task was presented using the Psychophysics Toolbox (Brainard, 1997) and projected onto a mirror screen above the participant’s eyes.

Participants were instructed that each face option was associated with a different probability of reward, that these probabilities could change slowly, and that their goal was to attempt to find the most rewarding option at a given time in order to earn the most money. They were also instructed that rewards were tied to the face identity and not the face position. Prior to the scanning session, participants completed a short practice version to familiarize them with the task and to ensure that their button responses reflected their intended choices.

Each of the options (S1-S4) was associated with a different probability of monetary reward. Across the 300 trials in the experiment, the reward probabilities diffused gradually according to Gaussian random walks, so as to encourage continual learning. Unbeknownst to the participants, to provide the opportunity of encoding
stimulus-stimulus relational structure, the faces were grouped into equivalent pairs (here referred to as faces S₁ & S₃ and S₂ & S₄). The chance of reward on choosing S₁ or S₃ (and similarly S₂ or S₄) was the same on any particular trial; however, trial feedback only displayed the reward outcome for the selected face. The reward probability for each pair of face stimuli changed over time, diffusing between 25% and 75% according to Gaussian random walks with reflecting boundary conditions. Two instantiations of two sets of random walks were generated, and these were then inverted (i.e., subtracting all probabilities from 100%) to give a total of four sequences (Figure 2.1B). So as to ensure that these strong positive correlations did not make the choice problem trivial (i.e., with all four options often having roughly the same value), a more modest negative correlation was included between the two sets of walks within each of these sequences (r² between pairs, -0.135 and -0.369; vs. r² = 1 within paired options). Reward probability sequences were counterbalanced between participants, as was the mapping of particular face stimuli to the underlying reward sequences.

After the completion of scanning, participants answered a series of questions that assessed their strategies during learning and their awareness of the contingencies across options. To further probe any knowledge of the underlying task structure provided by the equivalence relationships, participants were also given a questionnaire that included pictures of the four face stimuli. Participants were instructed to draw lines connecting the pairs of stimuli that for any reason seemed related to one another, and then to describe why they paired those options together (data available for 18 participants). Participants were then informed how much money they had won in the experiment.
Imaging procedure. Whole-brain imaging was conducted on a 3.0T Siemens Allegra head-only MRI system at NYU’s Center for Brain Imaging, using a Nova Medical NM-011 head coil. Head padding was used to minimize head motion; subsequent inspection showed that no participant’s motion exceeded 2mm in any direction from one volume acquisition to the next. Structural images were collected using a high-resolution T1-weighted MPRAGE pulse sequence (1 X 1 X 1 mm voxel size). Functional images were collected using a gradient echo T2*-weighted echoplanar (EPI) sequence with blood oxygenation level-dependent (BOLD) contrast (TR = 2000 ms, TE = 15 ms, flip angle = 82, 3 X 3 X 3 mm voxel size; 33 contiguous oblique-axial slices), tilted on a per-participant basis approximately 23° off of the AC-PC axis to optimize sensitivity to signal in the orbitofrontal cortex and the medial temporal lobe (Deichmann et al., 2003). The task was scanned in four blocks each of 310 volumes (10 min 20 s). For each functional scanning block, four discarded volumes were collected prior to the first trial to allow for magnetic field equilibration.

Behavioral Analysis. Model-based analyses were used to investigate participants’ learning and utilization of the reward equivalence structure to guide choices. Such analyses attempt to explain the timeseries of choices in terms of previous events, allowing precise, quantitative questions to be posed about the dynamics of behavioral adjustment. (See O’Doherty et al., 2007; Daw, 2011 for reviews of the methodology.)

First, we sought to test whether participants adjusted their choices dynamically in response to the rewarding outcomes. Because of the fluctuating probability of reward, we could not estimate a learning curve or a percent correct over the course of the task.
Instead, as in prior studies, a logistic regression model was fit to explain each participant’s sequence of choices in terms of two explanatory variables coding events from the previous trial: the choice made and whether it was rewarded (both coded as binary indicators) (Lau and Glimcher, 2005; Gershman et al., 2009; Daw et al., 2011; Li and Daw, 2011). In the present study the dependent variable is multinomial (i.e. choices over four options), so that the appropriate model is a conditional logit (McFadden, 1974), i.e. the link function is the softmax from reinforcement learning (Daw, 2011).

Having determined that participants’ choices were influenced by prior rewards, we next aimed to investigate more detailed aspects of learning using two variations of an RL model fit to the choice sequences (Sutton and Barto, 1998), as detailed below.

The model learns to assign an action value to each option, $Q_1 \ldots Q_4$, according to previously experienced rewards. These are assumed to be learned by a delta rule: if option $c$ was chosen and reward $r$ (1 or 0) received, then $Q_c$ is updated according to:

$$ Q_{c,t} = Q_{c,t-1} + \alpha*\delta_{c,t} \tag{1} $$

$$ \delta_{c,t} = r_t - Q_{c,t-1} \tag{2} $$

where the free parameter $\alpha$ controls the learning rate. To embody possible generalization of value across paired options with yoked drifting reward probabilities, the model includes a capacity to update the partner option yoked to the current choice. In particular, if option $c$ was chosen, with partner $p$, then in addition to updating the value of $c$, $Q_c$, as in Equations 1 and 2, $Q_p$ was also updated according to:
\[ Q_{p,t} = Q_{p,t-1} + \alpha_2 \cdot \delta_{p,t} \]  

(3)

\[ \delta_{p,t} = r_t - Q_{p,t-1} \]  

(4)

with the free parameter \( \alpha_2 \) controlling generalization learning rate. When \( \alpha_2 \) is set to zero, the model is blind to correlational structure, and corresponds to models studied previously (Daw et al., 2006; Schönberg et al., 2007; Gershman et al., 2009). In this sense, this no-generalization limit provides a null hypothesis or baseline model against which to test for generalization effects. With a non-zero generalization learning rate the model allows the reward feedback associated with a selected option (e.g., \( S_1 \) or \( S_2 \)) to update the value of its partner (\( S_3 \) or \( S_4 \), respectively). Because the models are otherwise identical, this parameter isolates generalization: i.e., we reasoned that if the model with a free generalization learning rate fit significantly better than the baseline one, then such a difference would be attributable to generalization across partners. Moreover, the estimated value of the learning rate measures the strength of the generalization effect (Daw and Shohamy, 2008). Note that a version of the model in which instead of moving partners’ values toward the obtained rewards, non-partners’ values are moved away from them (reflecting anti-generalization according to negative correlations; Hampton et al., 2006) makes predictions quantitatively similar to the version used here. This is because choice probabilities in the softmax (below) are driven only by the differences between \( Q \) values. Thus, for concreteness, and because the positive correlations were stronger in the reward schedules as programmed, we used the positively generalizing form of the rule.
Given value estimates on a particular trial, participants are assumed to choose between the options stochastically with probabilities \( P_1 \ldots P_4 \) according to a softmax distribution (Daw et al., 2006):

\[
P_{c,t} \propto \exp(\beta(Q_{c,t} + \varphi I(c_{c,t-1})))
\]  

(5)

The free parameter \( \beta \) represents the softmax inverse temperature, which controls the exclusivity with which choices are focused on the highest-valued option. The model also included a free parameter \( \varphi \), which, when multiplied by the indicator function \( I(c_{c,t-1}) \), defined as 1 if \( c \) is the same choice as that made on the previous trial, and zero otherwise, captures a tendency to choose (for positive \( \varphi \)) or avoid (for negative \( \varphi \)) the same option chosen on the preceding trial (Lau and Glimcher, 2005; Schönberg et al., 2007). Note that since the softmax is also the link function for the conditional logit model discussed above, this analysis also has the form of a regression from Q values onto choices (Lau and Glimcher, 2005; Daw, 2011) except here, rather than as linear effects, the past rewards enter via the recursive learning of Q, controlled, in nonlinear fashion, by the learning rate parameters.

In order to search for indications of generalization during the task (i.e. exploiting the relational structure underlying the gamble options), we compared the fit of two variants of the Q-learning model described by Equations 1-5: 1) the “base” model, where the generalization learning rate, \( \alpha_2 \), was set to zero, and 2) the “generalization” model, where \( \alpha_2 \) was a free parameter.
Although equivalence effects would be expected to evolve over time as participants gradually learned the equivalence, for simplicity and lacking a well supported formal model of the dynamics of such learning, we consider a simplified model in which $\alpha_2$ is taken as fixed across the experiment. Because the partner learning rate is thus fit to explain choices even over early parts of the task during which it is unlikely that participants will yet have detected any generalization structure, this is a conservative analysis in the sense that it will tend to underestimate the asymptotic equivalence effects (Daw and Shohamy, 2008).

For each participant, maximum likelihood values for the parameters $\alpha$, $\beta$, and $\phi$, as well as $\alpha_2$ for the generalization model, were estimated using a gradient search (repeated with 20 different starting points, decreasing the chance of local optima) over the likelihood of the participant’s observed choice sequence, for each trial conditional on the previous rewards and choices (Lau and Glimcher, 2005; Daw et al., 2006; Daw, 2011). In particular, log likelihood is computed as the sum over trials of $\log(P_c)$ for the actually chosen option using values learned by the model from the previously delivered rewards. A separate set of parameters was optimized for each participant.

To test whether the models provided a reliable account of participants’ behavior, we performed several analyses. First, we tested whether the base and full generalization models fit significantly better than chance (i.e. a model with no parameters, with $P_{c(s,t),t} = .25$ for all $t$), using likelihood ratio tests. The relative degree of improvement over the chance model provides a standardized descriptive index of how well a model fits, called pseudo-$R^2$ (Camerer and Ho, 1999; Daw et al., 2006), which we report for comparison with other studies. This is defined as $(R - L)/R$ where $L$ and $R$ are, respectively, the log
likelihood of the choices under the model (base or generalization) and under purely random choices \(P_{c(s,t),t} = 0.25\) for all \(t\).

For the critical comparison between models, the performance of the base model and the full generalization model were compared using likelihood ratio tests on the individual participants’ and summed group log likelihood values. Such a test examines the null hypothesis that any improvement in model fit is due to chance, correcting for the inclusion of additional free parameters (e.g., Stephan et al., 2009).

In order to reason about the prevalence of the two models across the population as a random effect that might vary across participants, we conducted an additional analysis using the Bayesian Model Selection (BMS) method of Stephan et al. (2009). In particular, we estimated Bayes factors (the posterior evidence for one model over the other; Kass and Raftery, 1995) using the AIC criterion (Akaike, 1974), and submitted these to the spm_BMS routine from SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK).

*Imaging analysis.* Preprocessing and data analysis was performed using Statistical Parametric Mapping software (SPM5; Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were realigned to correct for participant motion and then spatially normalized by estimating a warping to template space from each participant’s anatomical image (SPM5, “segment and normalize”) and applying the resulting transformation to the EPIs. Images were resampled to 2mm cubic voxels, smoothed with an 8mm FWHM Gaussian kernel, and filtered with a 128s high-pass filter.
For reinforcement learning model-based analysis of the fMRI data, we investigated correlations with trial-by-trial parametric signals derived from simulations of the model described above (Eq. 1-5). Data were analyzed using SPM5, under the assumptions of the general linear model. The events on each trial were modeled by half-second boxcar regressors at the time of stimulus onset and of outcome feedback. These two events were modulated by parametric regressors: the trial-by-trial probability of the chosen option (Eq. 5) on the stimulus onset, and the trial-by-trial prediction error (Eq. 4) on the outcome. Each event was also modulated by a second parametric regressor capturing the difference between probabilities or prediction errors in a model with and without generalization (formally, the partial derivative of the modeled quantity with respect to the partner learning rate; see below). Nuisance boxcar regressors were also included during the choice period (4 s) and outcome display periods (2 s) to account for general effects of visual stimulation.

To generate the parametric regressors for the imaging analysis, the free parameters for the learning model were chosen as follows. First, the learning model was re-estimated with the generalization learning rate, $\alpha_2$, set to zero. This was chosen so as to best characterize values and prediction errors under the null hypothesis of no generalization, allowing us to test (and perhaps reject) it at the neural level. Second, as has been noted previously (Daw et al., 2006), individual parametric fits in tasks and models of this sort tend to be noisy, and regularization of the behaviorally fit parameters across participants tends to improve a model’s subsequent fit to fMRI data. Accordingly (following previous work: Daw et al., 2006; Schönberg et al., 2007; Gershman et al., 2009), we generated regressors for each participant using a single setting of the RL
model’s free parameters, here taken as the mean over all participants of the best fitting individual estimates. The group means estimate the population-level parameters in a random-effects model of inter-subject variability (Holmes and Friston, 1998), and are thus a principled choice for the entire group. Note that although we thus do not characterize individual variability in most of the behavioral model parameters for the purpose of generating fMRI regressors, our approach does capture individual variability in the most important parameter for our questions of interest, the generalization learning rate, \( \alpha_2 \), since the prediction error partial derivative “difference” regressor capturing its effects in the fMRI model (see below) is taken as a random effect across subjects.

To investigate whether value-related neural signals reflect generalization of feedback across partner options, two additional regressors were included to accompany the base parametric RL regressors of reward prediction error and choice probability. These two “difference regressors” each characterize how one of these trial-by-trial parametric timeseries would change if the model included learning from partner option feedback (i.e., if the parameter \( \alpha_2 \) were nonzero). Intuitively, these regressors represent the difference between the probabilities (or prediction errors) generated according to two competing assumptions about the generalization learning rate \( \alpha_2 \): that it takes on some nonzero value \( \Delta \), vs. the null assumption that \( \alpha_2 = 0 \) (Wittmann et al., 2008; Daw, 2011; Daw et al., 2011). Thus, if the BOLD signal in an area is better correlated with the regressor timeseries for nonzero generalization (\( \alpha_2 = \Delta \)), then, given the additive nature of the GLM, the net BOLD signal will be best explained by a sum of contributions, from the main regressor (\( \alpha_2 = 0 \)) plus the difference regressor. If, instead, the BOLD signal correlates are best explained by \( \alpha_2 = 0 \), there should be no effect of the difference
regressor. In other words, this analysis separates a test for generic prediction error (without generalization) and an additional, orthogonal, test of whether such activity would actually be better explained by the prediction error including generalization (the test of the difference regressor in the same voxels). This approach (Wittmann et al., 2008; Daw et al., 2011; Bornstein and Daw, 2012) more cleanly separates these two inferences than simply including regressors generated according to both models and contrasting them (e.g., Hampton et al., 2008) particularly when the signals predicted by the models are correlated.

More formally, this additive approach approximates the (nonlinear) effect of an arbitrary \( \alpha_2 \) on the modeled probability or prediction error timeseries in the context of the standard linear analysis of the BOLD response by using a Taylor expansion of this nonlinear function around \( \alpha_2 = 0 \) and retaining the first-order (linear) term (Friston et al., 1998; Daw, 2011). This corresponds, in the above scheme, to taking the learning rate increment \( \Delta \) infinitesimally small, or equivalently, to defining the difference regressors as the partial derivatives of the modeled timeseries with respect to \( \alpha_2 \), evaluated at \( \alpha_2 = 0 \). Thus, if the BOLD response is better explained by a timeseries including nonzero generalization, the additive general linear model will explain the BOLD signal via the weighted sum of both regressors; in particular, a significantly positive effect will be estimated for the partial derivative. Voxels that show significant correlations with both prediction error and the prediction error difference regressor, or base chosen value and the chosen value difference regressor, exhibit activity that is better fit by a generalization learning model.
In order to test whether neural effects related to reinforcement learning variables were better explained by including effects of generalization, we identified activity correlated with basic RL variables, then tested for effects of the difference regressors (orthogonalized against the original variables to test only for residual activity), in the vicinity. To test whether these effects were significant in the same voxels (thus, whether activity in a voxel is best described by the weighted sum of both effects) we examined the conjunction of two tests, using SPM’s conjunction null (Nichols et al., 2005). Note that although the difference regressors were orthogonalized to the underlying prediction error variables, the validity of conjunction inference using the minimum t statistic does not depend on the conjoined tests being independent (Nichols et al., 2005).

Finally, we examined functional interactions between the striatum and the hippocampus during learning. We focused on a ventral striatum cluster identified in the above GLM as having a significant correlation with the prediction error difference regressor (6 mm spherical ROI; coordinates: 14, 8, -8). A psycho-physiological interaction (PPI) analysis was estimated to test for increases in functional correlation between the ventral striatum (the physiological variable) and other brain regions during choice trials (the psychological variable). The time course of activation from the ROI was extracted and deconvolved. This timecourse was interacted with the choice trial boxcar indicator and then convolved with the HRF. The model included the striatal timecourse by trial regressor, the trial regressor, and the unmodulated striatal timecourse regressor (Friston et al., 1997). We then correlated the resulting beta values with individual difference measures of the relative fit of the generalization model to behavior (calculated as the
difference between choice likelihoods for the base model vs. the generalization model (e.g. Hampton et al., 2008; Simon and Daw, 2011)).

fMRI model regressors were convolved with the canonical hemodynamic response function and entered into a general linear model (GLM) of each subject’s fMRI data. The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the GLM to account for residual effects of subject movement. Linear contrasts of the resulting SPMs were taken to a group-level (random-effects) analysis. We report results corrected for familywise error due to multiple comparisons using cluster size (Friston et al., 1993); this approach assesses the spatial extent of clusters defined by an initial and arbitrary uncorrected threshold, which we take as \( P < 0.005 \) for all analyses. Accordingly, for display purposes, we render all activations at this threshold. We conduct this correction either whole brain, or within small volumes for which we had an a priori hypothesis. In particular, in the striatum we used a hand-drawn mask of the right nucleus accumbens, based on prior studies showing robust prediction error and model-based influences in this region (Wittmann et al., 2008; Daw et al., 2011) (in both cases most robustly on the right). In the MTL we use an anatomically defined mask which included both the hippocampus and parahippocampus, derived from the AAL atlas (Tzourio-Mazoyer et al., 2002). All voxel locations are reported in Montreal Neurological Institute (MNI) coordinates, and results are displayed overlaid on the average of all participants’ normalized high-resolution structural images.
Results

Behavioral results

Over the course of the experiment, participants won $7.56 ± 0.10 (mean ± SEM across participants). Participants were able on most trials to enter a choice within the time constraints (9.6 ± 1.4 missed trials out of 300). On completed trials, response times were 1.16 s ± 0.02 (grand means ± SEMs across participants).

As the task provides only binary feedback about the selected option on each trial, information about similarities between options can only accumulate over multiple trials and switches between options. Because of these properties of the design, knowledge of the task structure may not often reach the level of explicit awareness. Participants shifted their choice selection an average of 115.10 ± 9.47 times (range 32-211), which provides an opportunity for participants to compare values across options. To investigate whether participants displayed explicit awareness of the relational structure of the task, after the experiment, we presented them with a display of the four options and asked them to indicate, by drawing connecting lines, which pairs of options seemed related in any way, and also asked them in a written follow-up question to describe any reasons underlying their answer.

Across the group, pairing performance did not differ from chance (33%; mean correct 22% ±10; data available for 18 participants), indicating that participants, collectively, were not explicitly aware of the manipulation. Individually, our criterion for explicit knowledge was both correctly pairing the options and exhibiting some explicit knowledge of the reward equivalence structure on the written question, a combination achieved by only one participant. (On the written question, that participant stated that “...
the pairs seemed to alternate when those 2 faces were ‘lucky.’”) These post-task measures suggest that the influence of the reward equivalence structure on choice behavior, as discussed below, is not likely due to participants’ explicit detection of the relational structure of the task.

**Reinforcement learning model of choices**

Next, we used the fit of computational models to examine the trial-by-trial dynamics of behavioral adjustment. In particular, such models allow quantifying how the choices depend on recent feedback, allowing questions to be asked about the specific nature of the updating: in particular, here, whether it reflected generalization between partners.

First, to examine whether participants adjusted their behavior dynamically to previous rewards, we fit a simple regression model measuring the extent to which each participant’s choice sequence was predicted by the reward on the previous trial, also controlling for the previous choice as an additional explanatory variable, as done previously (Gershman et al., 2009; Li and Daw, 2011). Consistent with prior reports, we found that across participants, the effect of the previous reward was significant (beta = 4.12 ± 1.12, t(20) = 3.67, P < 0.005), indicating participants learned choice preferences from previous rewards, while the effect of the previous choice was not significant (beta = 0.06 ± 0.26, t(20) = 0.22, P > 0.5).

Next, to examine whether this adjustment reflected the underlying hidden reward equivalence structure in the gambling task, we tested the fit of more detailed reinforcement learning models characterizing trial-by-trial adjustments in values for each
option. In particular, we compared models which differed only in whether they
generalized between partners, allowing us to test whether choice behavior evidenced any
generalization between equivalent options (S₁ & S₃ and S₂ & S₄) (Daw and Shohamy, 2008). Standard reinforcement learning models would assume that participants’ tendency
to choose an option is based on a learned value for that option which is updated only
from experience with outcomes from choices of that option. In contrast, a generalization
model embodies the idea that outcomes received for one option can influence learning
about the value of another option.

To address this question, we considered the fit of two different reinforcement
learning models. The “base” model consisted of a standard reinforcement learning model
blind to the relational structure of the task, while the “generalization” model extended the
base model to allow feedback about the present choice to update the value of the
unchosen partner option by way of an additional learning rate parameter. The two models
coincide when this additional parameter takes on the value zero. A similar generalization
model has been shown to better fit participant choice behavior in a prior study that
reported the results of a task analogous to the current one (Daw and Shohamy, 2008).

First, we confirmed that both the base and generalization models each explained
choices better than chance. This was the case both in the aggregate over participants
(likelihood ratio tests; $\chi^2_{63} = 6,694.20$; $\chi^2_{84} = 6,833.10$; P’s < 1e-16) and also individually
for all participants for both the base and the generalization model at P < 0.0001. Pseudo-
r² statistics (a descriptive measure of model fit appropriate for comparing between
studies) were 0.38 ± 0.17 for the base model and 0.39 ± 0.17 for the generalization
model.
Next, we compared the two models’ fits to one another to determine whether there was evidence for generalization. For choice likelihoods aggregated over all participants (equivalent to assuming all participants complied with one model or the other, and testing which one), the difference in log choice likelihoods (Table 2.1) was 69.4 in favor of the generalization model, i.e. the choices were \(\exp(69.4)\) more likely given the generalization model than the base model (Kass and Raftery, 1995). We formally tested whether such an improvement was expected due to chance given the extra free parameters with a likelihood ratio test; the restriction to the base model was indeed rejected in favor of the generalization model (likelihood ratio test, \(c^2_{21} = 138.86; p<1e^{-16}\); Table 2.1).

<table>
<thead>
<tr>
<th>Model</th>
<th>-LL</th>
<th>Aggregate LR Test</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>(\varphi)</th>
<th>(\alpha_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>5386.5</td>
<td>---</td>
<td>0.68 ± 0.06</td>
<td>4.50 ± 1.09</td>
<td>0.19 ± 0.04</td>
<td>---</td>
</tr>
<tr>
<td>Generalization</td>
<td>5317.1</td>
<td>(c^2_{21} = 138.9)</td>
<td>0.69 ± 0.06</td>
<td>4.86 ± 1.11</td>
<td>0.19 ± 0.03</td>
<td>0.09 ± 0.04</td>
</tr>
</tbody>
</table>

Table 2.1. Reinforcement learning model fits.

Reinforcement learning model fits. For comparing the base model and the generalization model, which incorporates value generalization across partnered options, shown are negative log-likelihood (-LL), aggregate likelihood ratio test statistic \((c^2)\), and random-effects maximum-likelihood parameter estimates (mean ±SEM across participants) for both the base and generalization reinforcement learning models.

The foregoing analyses aggregated evidence across participants. We next sought to address whether there were individual differences and whether the effects might be driven by outliers. Examining individuals, likelihood ratio tests also rejected the base model for 11 of 21 participants considered individually (at \(P < 0.05\); 12/21 at \(P < 0.06\)).
In order to more formally examine evidence for either model at the group level, allowing for the possibility that the existence of generalization might vary across participants (i.e. taking the identity of the best-fitting model as a random effect), we conducted an additional Bayesian analysis of the choice fits, fitting a hierarchical model in which participants are assumed to be drawn from either sort and estimating the proportions (Stephan et al., 2009). The estimated fraction of generalizers in the population was 0.853 (compared to 0.147 for non-generalizers); the “exceedance probability”, or posterior probability that the generalization model was the more prevalent of the two, was 99.9%.

The hypothesis of generalization may also be assessed at the group level by treating the learning rate controlling generalization as a random effect analogous to population-level effects in fMRI (Holmes and Friston, 1998). Across participants, the best fitting estimates were indeed significantly different from zero ($t_{(20)} = 2.40, P < 0.05$; range -0.01 to 0.69, Table 1; note that to render this test meaningful it is important that we did not constrain the estimated parameter to be positive).

Though significant, the generalization effect was modest in size: on the average over participants, generalization learning rates were approximately 13% of the primary learning rate. We might expect generalization to be fractional, due to participants’ potentially incomplete detection of the relationship. In particular, our model likely underestimates the asymptotic degree of generalization, since for simplicity it treats the parameter as constant throughout the experiment (see Methods), in effect averaging over early parts of the experiment in which the relationship could yet not have been learned. Absent a well supported quantitative model of the timecourse of such learning, we separately estimated generalization learning rates for the first and second half of the
experiment. The estimated generalization learning rates were significantly greater in the second half (first half: 0.059 ± 0.025; second half: 0.154 ± 0.046, $P < 0.05$ one-tailed reflecting the directional hypothesis), which suggests that our model is detecting the expected increase in generalization knowledge over the course of the experiment.

Intriguingly, the single participant that displayed clear evidence of being explicitly aware of the generalization task structure showed the greatest model likelihood benefit for the generalization model and the second-highest fit generalization learning rate. Importantly, however, excluding this participant from the group likelihood ratio test, Bayesian model selection analysis, and parametric tests did not affect the significance of the results. This suggests that while most participants were not explicitly aware of the generalization structure, our generalization model clearly detected the single participant that did exhibit awareness as an outlier, supporting the validity and sensitivity of our approach.

Together, these results provide evidence that participants utilized the underlying relational reward equivalence structure to generalize reward feedback across equivalent options and guide their choice behavior.

**Imaging results**

Our analyses of the behavioral data established that the generalization reinforcement learning model, which embodied the generalization of value across pairs of equivalent options, provided a better fit to participants’ choice behavior than a reinforcement learning model blind to this relational structure. Thus, we turned to the BOLD fMRI data to investigate neural correlates of this generalization knowledge. In
particular, we sought activity correlated with reward predictions and prediction errors as produced by simulations of the reinforcement learning model under the null assumption of no generalization, and then tested whether this activity showed additional evidence of generalization knowledge. We particularly sought to test whether BOLD correlates of reward prediction error in the ventral striatum were naïve to generalization, as would be predicted under the standard model of these responses, and if these signals originate in a procedural learning system entirely separate from a putative cortico-hippocampal system capable of detecting relations and generalizing from them (Daw, 2011; Daw et al., 2011).

We focused first on activity correlated with the reward prediction error when the outcome is revealed. Since reward prediction errors report the difference between received and expected rewards, they may reflect the effects of generalization (if any) on the expectations. In particular, if outcomes received for some option also affect the value predicted for its partner, they will affect the prediction error reported on subsequent choices of the partner option. In contrast, such generalization-driven updating of values for an unselected option is not possible in standard stimulus-reward association learning models. To distinguish these possibilities, the fMRI model included a parametric regressor for the base prediction error, assuming no generalization, and a second “difference regressor” (technically, the partial derivative of the error signal with respect to the generalization learning rate, or equivalently the difference between the signals predicted by models with and without small amounts of generalization), characterizing how it would be expected to change if generalization were included (see Methods). In particular, the sum of the base prediction error and difference regressors, in any weighted combination, corresponds approximately to the prediction error from a model including
generalization (Figure 2.1B). Thus, since the general linear model used for fMRI analysis is additive, if BOLD responses in a region significantly reflect effects of both regressors, then the net activity there is better explained by a prediction error including generalization, and the region may support the value generalization effect we observed in participants’ behavior.

The difference regressor for prediction error across participants included a mean number of 105.3 ±7.0 positive deflections and 173.4 ±7.4 negative deflections. Difference regressor values were often most extreme when participants switched choices, as this is when the generalization model makes the most divergent predictions from the base model. To illustrate this effect, consider the case where an option (e.g. S1) has been rewarded on the last several trials, but the participant switches to choosing the partnered option on the next trial (e.g. S3). In the generalization model but not the base model, the value for S3 has increased, and this expectation will modulate the prediction error signal. Here, this will lead the difference regressor to include a negative deviation: if the choice is rewarded, this is less of a positive “surprise” to the generalization model, while if it is not rewarded, this omission is more of a negative surprise.

Accordingly, we first localized regions where BOLD activity correlated with prediction errors derived from the base reinforcement learning model. Reward prediction error correlates have been found most prominently in the ventral striatum (Knutson et al., 2001; Pagnoni et al., 2002; McClure et al., 2003; O'Doherty et al., 2003; Delgado et al., 2005; Daw et al., 2006; Lohrenz et al., 2007; Schönberg et al., 2007; Hare et al., 2008), a region densely innervated by midbrain dopamine neurons (Falck and Hillarp, 1959; Knutson and Gibbs, 2007). Replicating these findings, in the current experiment,
prediction error at reward outcome correlated with BOLD responses throughout the bilateral ventral striatum (Figure 2.2, left; right peak $z = 5.57$ (14, 4, -14), left peak $z = 5.27$ (-22, -4, -16); both clusters were significant whole-brain FWE-corrected for cluster size).

We next examined whether residual activity in this region reflected effects that could be explained by generalization of value between options. Indeed, activation in a region of the right ventral striatum significantly correlated with the difference regressor designed to capture the effects of generalization on prediction error (Figure 2.2, center; $z = 3.16$ (14, 8, -8), $P < 0.001$ uncorrected; $P < 0.01$ SVC for FWE in an a priori right nucleus accumbens anatomical ROI). A conjunction analysis (Figure 2.2, right; $P < 0.001$ uncorrected; $P < 0.01$ SVC) verified that this effect was spatially overlapping with the prediction error itself and therefore (see Methods) that the net activity in this region was better explained by a prediction from the generalization model. (A similar sub-threshold cluster was observed in the left ventral striatum.)

Figure 2.2. Ventral striatum BOLD and value generalization.

Ventral striatum BOLD signals are best described by a model that incorporates generalization knowledge. A, Prediction error, left. Prediction error difference due to value generalization, middle. B, Conjunction of prediction error and prediction error difference due to generalization ($P < 0.05$, SVC; all $P < 0.005$ unc., for visualization).
Thus, in contrast to predictions based on simple reinforcement learning models, the net BOLD signal in the right ventral striatum, a region often characterized by a reward prediction error response, is best explained by a reinforcement learning model that incorporates generalization knowledge. This result is consistent with other recent indications that the striatal error signal is more sophisticated than previously suspected (Daw et al., 2011; Simon and Daw, 2011).

Next, we asked whether generalization knowledge was also reflected in anticipatory value-related signals during the choice period. Again, we first localized regions where activity correlated with the value of the selected option (the “chosen value”) during the choice period. Here, following evidence from unit recordings that action values in the brain are normalized between options (Platt and Glimcher, 1999; Dorris and Glimcher, 2004; Sugrue et al., 2004), and previous fMRI work (Daw et al., 2006), we define an action’s value by the probability that the model predicts it will be chosen, which (Equation 5 in Methods) is a normalized transform of the raw value. Prior reinforcement learning studies of learning and decision making have often found correlates of chosen value in the ventromedial prefrontal cortex (Daw et al., 2006; Kim et al., 2006; Plassmann et al., 2007; Hare et al., 2008; Boorman et al., 2009; Gershman et al., 2009; Palminteri et al., 2009; Smith et al., 2010). Accordingly, at a reduced whole-brain threshold, we also observed a cluster of activation in the ventromedial prefrontal cortex correlated with value \( (z = 2.97 (-6, 58, -18), P < 0.001 \text{ unc}) \). The most extensive region of correlation, however, was observed bilaterally in the hippocampus (Figure 2.3; right peak \( z = 3.88 (34, -6, -26) \), left peak \( z = 3.49 (-16, -22, -20) \); \( P < 0.05 \text{ cluster-corrected in a medial temporal lobe mask} \)). Chosen value correlates in the hippocampus
have not often been reported in previous studies of reward learning and decision making, but this finding is consistent with several more recent reports of value encoding in the hippocampus in categorization and passive viewing tasks (Kumaran et al., 2009; Lebreton et al., 2009; Dickerson et al., 2011).

Figure 2.3. Hippocampal BOLD correlated with value.

Hippocampal activation correlated with chosen value during the choice period of the reward equivalence task. (P < 0.05, SVC; P < 0.005 unc., for visualization.)

We then asked whether these responses also reflected generalization knowledge. An analysis of the value difference regressor did not show any significant correlation in the ventromedial prefrontal cortex. In the left hippocampus, we observed a cluster correlated with difference regressor at an uncorrected whole-brain threshold (P < 0.005), but this activation did not survive cluster-correction based on a medial temporal lobe mask. The lack of significant evidence for generalization effects in these signals is unexpected in light of our hypothesis that the hippocampal system might support the generalization.
To examine this hypothesis further, we turned to individual differences in generalization as expressed behaviorally and tested their relationship to functional connectivity between the striatum and hippocampus. To assess connectivity, we conducted a psycho-physiological interaction (PPI) analysis using as a seed the region of right ventral striatum showing a significant fit to the prediction error difference due to generalization. Overall, trial-related activity in the ventral striatum was significantly correlated with activity in widespread brain regions, including multiple clusters in the hippocampus. We correlated the degree of connectivity from ventral striatum, across participants, with the model fit benefit provided by the generalization model to the choice behavior (the difference in choice likelihoods between the generalization and null models; n = 20, excluding a single outlier whose benefit was > 2 SD from the mean (the participant who exhibited high awareness of task structure)). We found that the degree of striatal-hippocampal connectivity was significantly predicted by the generalization model improvement in fit to choices ($z = 4.0 \ (26, -18, -16)$; Figure 2.4). Further, the cluster showing a connection between connectivity and generalization overlapped with the regions of the hippocampus where the BOLD signal exhibited a significant correlation with choice value. This result is consistent with the hypothesis that the hippocampus (and more specifically, hippocampal-striatal functional connectivity) contributes to choices that benefit from generalization across correlated options.
Figure 2.4. Ventral striatum-hippocampus connectivity related to generalization.

Psychophysiological interaction (PPI) between task and ventral striatal activity is predicted by the degree that a participant’s choice behavior is better fit by the generalization RL model. (n = 20; FWE SVC in the MTL at P < 0.01) (images thresholded at P < 0.005, unc. for display).
Discussion

Our data show that choice behavior and feedback-related BOLD signals in the striatum are both influenced by the generalization of reward across equivalent options, as revealed by a novel reinforcement learning task in which payoff probabilities between pairs of options were correlated, providing an opportunity for participants to encode stimulus-stimulus relations. This structure, wherein individual cues have common outcomes, leading to generalization between them, is conceptually similar to the structure of acquired equivalence tasks used in research on memory to probe hippocampal representations (Myers et al., 2003; Daw and Shohamy, 2008; Shohamy and Wagner, 2008). However, here the common outcomes are likelihood of reward, rather than (as in paired associate learning used in human studies) the identity of the outcome stimulus. Nevertheless, we found that the influence of this shared reward probability on both choice behavior and striatal BOLD signaling was captured by a reinforcement learning model that, whenever feedback was received about an option’s value, also fractionally updated the value of its equivalent partner. Notably, such generalization on the basis of correlational structure is not predicted by standard reinforcement learning models commonly used to describe reward-driven learning and associated neural responses in the midbrain dopamine system.

The present results suggest that human participants do indeed encode structure and generalize across correlated choice options during RL. One ambiguity that remains is to what extent our effects are driven by positive correlations between “equivalent” options or, instead or additionally, by weaker negative correlations that were also included between the two equivalent pairs in our reward schedules. For concreteness (and
because the positive correlations were objectively much stronger), our analysis assumed positive generalization. However, in our model and overall framework (see Methods) generalization driven by correlations could, in principle, also arise due to negative generalization between anti-correlated non-partners. Both conceptually and mathematically (due to symmetries in the softmax choice equations), positive and negative generalization might be expected to have quantitatively quite similar effects on choices and BOLD signals. Therefore, disentangling the relative contributions of similarity and distinctiveness to generalization awaits further experiments manipulating the positive and negative correlations independently. In RL tasks, unambiguous negative generalization between options has been observed when their values are strongly anti-correlated due to a serial reversal contingency (Hampton et al., 2006; Bromberg-Martin et al., 2010b). Importantly, the central cognitive and computational issues and our basic conclusions about generalization according to structure crosscut this distinction between positive and negative generalization.

_Hippocampus and value_

On the basis of the analogy between the correlational structure embedded in our task and that in acquired equivalence studies (Coutureau et al., 2002; Myers et al., 2003; Shohamy and Wagner, 2008), we hypothesized that learning of correlational structure would implicate the hippocampus. Our data provide somewhat mixed support for this hypothesis. The most direct evidence in favor was our finding that connectivity between the striatum and hippocampus predicted the degree to which participants’ choice behavior was better described by the generalization model.
Also consistent with hippocampal involvement in this task, we found strong and widespread covariation of the BOLD signal in bilateral hippocampus with chosen option value, derived from the reinforcement learning model. This activation stands out in the context of the literature on reinforcement learning tasks similar to ours, particularly since similar activity is much more widely reported in ventromedial PFC (Daw et al., 2006; Kim et al., 2006; Plassmann et al., 2007; Hare et al., 2008; Boorman et al., 2009; Gershman et al., 2009; Palminteri et al., 2009; Smith et al., 2010), where value-related activity was relatively modest in the present study.

An intriguing possibility is that the inclusion of structure in the present task recruited systems for valuation at least partly distinct from those exercised by other tasks. This hypothesis is consistent with other recent reports of hippocampal activation related in some way to stimulus value, which used task designs (active learning, passive observation, or model-based reinforcement learning) that might enhance the relevance of relational information relative to standard reinforcement learning tasks (Kumaran et al., 2009; Lebreton et al., 2009; Dickerson et al., 2011; Simon and Daw, 2011). However, future studies will be necessary to directly test this hypothesis by specifically comparing learning with a relational component vs. without within a single study.

We were unable to demonstrate the effects of value generalization quantitatively in hippocampal correlates of value, even though effects of generalization were visible in the striatum. Based on our hypothesis that the hippocampus supports generalization between options’ values, this result is puzzling and it may indicate that the hypothesis was incorrect. At the same time, this null result should not be over-interpreted; this may be due, for instance, to our less refined quantitative characterization of the neural
correlates of chosen value in the hippocampus, relative to prediction errors as studied in the striatum. In particular, it has been persistently unclear whether neural activity in many different parts of the brain correlates with chosen value linearly, or better via some nonlinear transform or normalization such as the softmax employed here (Platt and Glimcher, 1999; Corrado et al., 2005; Daw and Doya, 2006; Daw et al., 2006). However, the form of this relationship is not well specified, and because our analysis seeking neural correlates of generalization is based on a linear approximation (a first-order Taylor expansion of the modeled signal’s dependence on the learning rate for generalization), it is likely particularly sensitive to any misspecification of this sort. (See also Daw et al.’s, 2011, discussion of value-related BOLD activity in ventromedial PFC vs. striatum for a similarly equivocal result from a similar analysis.)

Because our results give mixed support to the hypothesized role of the hippocampus in value generalization, it is worth considering whether our design changes some key aspects of acquired equivalence that engage the hippocampus. The chief difference from most acquired equivalence studies is that in our task equivalence is driven by value (e.g. stimulus-reward equivalences) rather than arbitrary stimulus-stimulus associations of the sort used in prior human and animal acquired equivalence studies (Coutureau et al., 2002; Myers et al., 2003; Shohamy and Wagner, 2008). However, acquired equivalence has also been demonstrated via value in rodents (albeit without neural manipulations to test hippocampal involvement; Honey and Hall, 1989). Moreover, entorhinal lesions have been shown in rodents to affect acquired equivalence using a task that counterbalances both stimulus-stimulus and stimulus-reward outcomes between partners and non-partners (Coutureau et al., 2002). This suggests that the
medial-temporal lobe memory system may be implicated more generally in encoding and inferring equivalence, rather than specifically for stimulus-stimulus encoding.

Finally, although conscious awareness is sometimes viewed as a characteristic of hippocampal episodic representations, the finding that most participants in our task did not report awareness of the relational structure does not preclude hippocampal involvement. Work in both memory and decision making isolates different types of representations operationally by the nature of the information coded rather than by self-report; in this context, there is much evidence of hippocampal involvement in relational coding absent conscious awareness (Greene et al., 2006; Shohamy and Wagner, 2008; Hannula and Ranganath, 2009). Nonetheless, it is interesting to note that in the current study the single participant who showed clear evidence of awareness of the task structure also showed the strongest evidence of generalization in trial-by-trial choices. Future studies are necessary to more directly probe the role of awareness of structure on generalization and on the role of the hippocampus and the striatum in generalization-guided choices.

**Ventral striatum and value generalization**

Although our task elicited relational coding not typically implicated in reinforcement learning tasks and may have recruited additional neural circuitry subserving this function, we nevertheless also observed the now-standard correlates of reward prediction error in the ventral striatum (Knutson et al., 2001; McClure et al., 2003; O'Doherty et al., 2003; Delgado et al., 2005; Lohrenz et al., 2007; Hare et al., 2008). However, here the net striatal activation was better explained by error signals from the augmented model that
learned its predictions about an option’s rewards not just from feedback about that option, but also by generalizing from its partner. By design, such a finding goes beyond what can be explained by standard reinforcement learning models without such augmentation, and demonstrates that the ventral striatum has access to information about correlational structure of a sort that goes beyond the simple, stimulus-reward learning normally associated with this area. The question of whether striatal value signals reflect such generalization was left open by a related study by Hampton et al. (Hampton et al., 2006), who investigated generalization in a serial reversal task (which causes two options’ values to be negatively correlated, rather than positively, as here). There, value correlates in ventromedial PFC were shown to reflect generalization, but the same question was not asked about prediction errors in the striatum. (Also, unlike the present study, participants in the Hampton task were instructed as to the reversal contingency.)

The finding of generalization in the striatal error signal also cuts against two-system accounts of both reinforcement learning and of memory systems, which envision that a standard temporal-difference learning system is responsible for limited, “habitual” behaviors, whereas more sophisticated decision-making phenomena drawing on cognitive maps or action-outcome associations (in memory terms, relational representations) are segregated in a parallel, competing network for “model-based” reinforcement learning (Doya, 1999; Daw et al., 2005; Balleine et al., 2008; Rangel et al., 2008; Redish et al., 2008). Contrary to our results, such an architecture predicts that signals originating within the putative temporal-difference system (notably, the ventral striatal prediction error) will be naïve to relational information even when behavior, under the control of the more sophisticated system, reflects it.
Two other recent results attempting to interrogate the model-free vs. model-based distinction more explicitly also found evidence for model-based effects on striatal prediction error signals (Daw et al., 2011; Simon and Daw, 2011). Altogether, these results suggest that the systems are more interacting than separate, an idea even more directly supported by the present study’s results regarding functional connectivity between striatum and hippocampus. That said, another possibility regarding the present dataset is that generalization effects do not arise from a full model-based planning system, but rather, from standard temporal-difference learning operating over an input representation that reflects the relationship between the options (i.e., which maps options to values but with equivalent options coded in an overlapping fashion; Gluck and Myers, 1993; Moustafa et al., 2009). Such an interpretation is also consistent with recent evidence from a two-phase acquired equivalence task that suggested that generalization effects arose already during the initial learning phase rather than via inference about equivalent relationships conducted during the probe phase (Shohamy and Wagner, 2008) (as would be expected from a model-based reinforcement learning system).

Although some results suggest that prediction errors in striatal BOLD may in part reflect dopaminergic inputs there (Pessiglione et al., 2006; Knutson and Gibbs, 2007; Schott et al., 2008; Schönberg et al., 2010), it is not possible to isolate the underlying neural cause for our effect, or in particular to conclude whether prediction errors carried by dopamine neurons also similarly reflect generalization. A related point is that the net BOLD signal in an area likely superimposes multiple underlying neural causes – including local processing and activity from different inputs. Thus, although our analysis uses the conjunction of multiple additive effects to assess what sort of prediction error
signal best explains the net BOLD response, it is not possible to exclude the possibility that these effects have different neural sources, and in particular that the generalization-related activity originates from a different source than the prediction error. All these questions could best be answered using unit recordings. However, in this respect it is interesting that our results are strongly reminiscent of a recent neurophysiological study in nonhuman primates, which showed that dopamine neurons also reflect values learned by generalization between two (negatively correlated) options in a serial reversal task (Bromberg-Martin et al., 2010b).

All these results (but not the idea of strictly segregated learning systems) are broadly consistent with strong anatomical connections between the hippocampus and the mesolimbic dopamine system. Intriguingly, in the present dataset, we find that functional connectivity between these regions, the ventral striatum and hippocampus, is predicted by the degree that participant’s choices were fit by the generalization model. Anatomically, the ventral striatum may gain access to relational representations via direct projections there from the hippocampus and medial temporal lobe (Kelley and Domesick, 1982; Cohen et al., 2009). Conversely, value information in the hippocampus may arrive via significant projections from midbrain dopaminergic neurons of the ventral tegmental area (Dahlström and Fuxe, 1964; Swanson, 1982; Frey et al., 1990; Gasbarri et al., 1994b; Huang and Kandel, 1995; Otmakhova and Lisman, 1996). These latter connections have broader implications for how hippocampal memories are influenced by reward, motivation and predictions (e.g. Adcock et al., 2006; Shohamy and Wagner, 2008; Kuhl et al., 2010; see Shohamy and Adcock, 2010 for review).
Limitations and future directions

One limitation of the present study is that, although our findings demonstrate that participants used the equivalence between the options to guide choices and that this effect increases in the second half of the experiment, our reinforcement learning model does not explicitly characterize the learning of the equivalence. In order to focus on the question of whether participants’ value learning reflected the equivalence structure, we took the degree of such learning and the underlying equivalence structure over which it operated as fixed throughout the task. For the questions of the present study, the main consequence of this approach is likely to underestimate the asymptotic size of the generalization effect, but it leaves open the question of how learning of the equivalence structure occurred. Accounts of such learning are reasonably well understood (at least in the abstract, it can be accomplished by Bayesian model comparison; Griffiths and Tenenbaum, 2005; Courville et al., 2006; Kemp and Tenenbaum, 2008); however, the present experimental design is not well suited to testing them. In particular, since the actual equivalence structure was fixed throughout the task, the learning of it occurred alongside many other potentially confounding changes (e.g. representational, strategic, or habituation) that may occur simply with time on task; a more targeted design would incorporate dynamic equivalencies so as to test different dynamic accounts of how participants follow them.

In general, our results highlight the promise of integrated investigations of memory and decision making. While often studied separately, it is clear that memory, if is it to be behaviorally beneficial, exists to guide decisions (Buckner, 2010; Shohamy and Adcock, 2010). A growing number of studies already focus on the cognitive and neural underpinnings of the use of different types of information in decision making (Johnson et
al., 2007; Daw and Shohamy, 2008; Shohamy and Adcock, 2010; van der Meer et al., 2010). Future studies may further probe how and when these different types of memory are reassembled into behavior by studying more complex decision processes and environments in conjunction with computational models (e.g. Daw et al., 2005). In this respect, our data point to the ability of the striatum to utilize information characteristic of relational memory systems, thus suggesting at least one underexplored way in which past experience can drive future choices.
Chapter 3: Preference by Association: A Neural Mechanism for How Memory Biases Decisions


Because of journal requirements, formatting for this chapter is includes the introduction, results, and discussion all as part of the main text.
In Chapter 2, I demonstrated that higher-order relational structure in a reward learning game impacts choice behavior and striatal prediction error signaling. Building on these findings, in Chapter 3 I explore the impact of relational associations on decision making in an experiment designed to require an interaction between relational and reward learning.

Abstract

Every day people make new choices between alternatives that they have never directly experienced. Yet, such decisions are often made rapidly and confidently. Here we show that the hippocampus, traditionally known for its role in building long-term declarative memories, enables the spread of value across memories, thereby guiding decisions between new choice options. Using functional brain imaging in humans, we discovered that giving people monetary rewards led to activation of a pre-established network of memories, spreading the positive value of reward to non-rewarded items stored in memory. Later, people were biased to choose these non-rewarded items. This decision bias was predicted by activity in the hippocampus, by reactivation of associated memories, and by connectivity between memory and reward regions in the brain. These findings explain how choices among new alternatives emerge automatically from the associative mechanisms by which the brain builds memories. Further, our findings demonstrate a novel role for the hippocampus in value-based decisions.
Main Text

Decisions are sometimes guided by direct past experience: If a choice led to a good outcome in the past, people are likely to make that same choice again. This process is known to depend on reward learning mechanisms in the striatum (Schultz, 2006; Rangel et al., 2008). But frequently in life we have to decide between options we have never considered before. It has been suggested that such decisions could be guided by associative memory (Brogden, 1939; Tolman, 1948; Eichenbaum, 2000); however, surprisingly little is known about how this happens.

We investigated the mechanism by which neural circuits for memory modulate value and guide decisions about new choice options. Our central hypothesis was that the hippocampus enables the positive value of reward to spread across associated memories, thereby increasing the value of items that were never rewarded. Specifically, we hypothesized that receiving reward can lead to two simultaneous and interactive processes: the direct learning of stimulus-reward associations in the striatum, and the spread of reward to associated items stored in memory via the hippocampus. This hypothesis is grounded in two essential features of how the hippocampus builds memories. First, the hippocampus encodes relationships between items and events that appear together, forming an associative link between them (Eichenbaum and Cohen, 2001; Hannula and Ranganath, 2009; Staresina and Davachi, 2009). Second, because of this associative link, when a person later encounters one item, the hippocampus can complete the pattern and automatically reactivate the neural representation of the other item (Polyn et al., 2005; Foster and Wilson, 2006; Kuhl et al., 2010; Kuhl et al., 2011), allowing the integration of old memories with new ones.
We reasoned that these features of memory formation in the hippocampus could provide a mechanism by which reward experiences can systematically change the value of items that were never rewarded – they gain a positive value merely by association. If so, then this mechanism predicts that later, when confronted with a decision, people will be biased to choose items that were never rewarded in the past (Brogden, 1939; Kimmel, 1977; Walther, 2002; Wimmer and Shohamy, 2011). By emphasizing the associative nature of processes in the hippocampus, regardless of awareness (Eichenbaum and Cohen, 2001; Hannula and Ranganath, 2009; Turk-Browne et al., 2009; Schapiro et al., 2012), this mechanism further predicts that the hippocampus might bias value even when associations are not explicitly remembered.

Figure 3.1. Decision bias task.

The task consists of three phases: Association learning, Reward learning, and Decision making. A, In the Association phase, participants were exposed to a series of pairs of pictures (S₁ and S₂ stimuli), while performing a cover task to detect “target” upside-down pictures. S₁ stimuli were either face, scene, or body part pictures. S₂ stimuli were circle images. B, In the Reward phase participants learned through classical conditioning that half...
of the $S_2$ stimuli were followed by a monetary reward ($S_2^+$), while the other $S_2$ stimuli were followed by a neutral outcome (no reward, $S_2^-$). $S_1$ stimuli never appeared in this stage. In the Decision phase, participants were asked to decide between two stimuli (both $S_1$ or both $S_2$) for a possible monetary win; no feedback was provided and all gains were awarded at the end of the experiment. Decision bias was operationalized as the tendency to choose $S_1^+$ over $S_1^-$ stimuli in this phase.

To test our prediction that reward will spread across associated memories, we used functional magnetic resonance imaging (fMRI) to measure brain responses during a learning and decision task designed to test how associative memory biases decisions between new choice options (Wimmer and Shohamy, 2011) (**Figure 3.1**). First, we had participants ($n = 28$) build new associative memories by exposing them to regularities between pairs of neutral stimuli (denoted here as $S_1$ and $S_2$). These regularities were encoded incidentally while participants performed a cover task (**Figure 3.1A**).

Next, we associated value with some items using a classical conditioning paradigm in which half of the $S_2$ stimuli ($S_2^+$) were now followed by a monetary reward (**Figure 3.1B**). This procedure is known to enhance the value of the directly rewarded items via well-described reward learning mechanisms in the striatum (Rangel et al., 2008). Critically, we hypothesized that at the same time, associative memory processes in the hippocampus would activate the specific $S_1$ items associated with the rewarded $S_2$ items, resulting in a transfer of the reward to the $S_1$ items as well, through hippocampal-striatal connectivity (Shohamy and Adcock, 2010; Pennartz et al., 2011). This would increase the value of $S_1$ items that were linked to the rewarded $S_2$ items ($S_1^+$), creating a bias towards choosing these items in the future – despite the fact that these items were never rewarded and were not even seen during reward learning.
To measure the effect of associative memory on value, in the final phase participants made a series of decisions in which they had to choose between two $S_1$ items, selecting the ‘luckier’ one for potential winnings, awarded at the end of the experiment (Figure 3.1C, top). Absent any spread of reward, participants should be equally likely to choose any of these non-rewarded items and brain activity during the prior Reward phase should be unrelated to these decisions. However, if reward spreads and biases decisions, then participants should be biased towards choosing those non-rewarded $S_1$ items that were previously associated with the $S_2$ rewarded items. Thus, we operationalized “decision bias” as the tendency to chose $S_1+$ items over $S_1-$ items, and we hypothesized that decision bias should be related to neural processes in the hippocampus during reward learning.

To test this hypothesis, we focused our analyses on brain activity during the Reward phase and asked whether activity during this phase was related to later biases in decisions. We made three specific predictions about the neural mechanisms giving rise to the spread of value to new choice options: (1) If associative memory processes underlie shifts in value, then biased decisions should be predicted by the magnitude of activation in the hippocampus during reward learning; (2) If reactivation of associations is the mechanism by which value spreads, then during reward learning there should be evidence of neural reactivation in visual areas that represent the original $S_1$ items; (3) If decision bias stems from interactions between associative memory processes in the hippocampus and reward learning processes in the striatum, then bias should be related to functional connectivity between these two regions.
Behaviorally, participants tended to choose the $S_2^+$ over the $S_2^-$ items in the Decision phase, indicating successful reward learning. Interestingly, decision bias in favor of $S_1^+$ items varied markedly both within and across individuals: Most participants displayed a bias in favor of $S_1^+$ items, but some did not (Figure 3.2A). Within individuals, this measure of bias was strong for some associations, but weaker for others. This variability in behavior allowed us to ask: What are the neural mechanisms that support decision bias?

Figure 3.2. Decision bias behavior and bias-predicting hippocampal activity.

Decision bias varies within and across participants and is related to activation in the hippocampus during the Reward phase. A, Decision phase preferences for $S_2^+$ stimuli (grey) and $S_1^+$ stimuli (average, blue; within-participant mean, black; error bars represent $\pm$ SEM). B, Decision bias variability, within-participant. Each participant was exposed to three different types of stimuli, allowing for differential bias across associations. Individual mean decision bias points from 2A were divided, ranked by level of bias, and averaged across participants. C, Hippocampal activation within-participants during the Reward phase ($S_2$ stimulus presentation and outcome) predicted the subsequent level of bias for associated $S_1$ stimuli [$z = 4.00 (26, -34, -12); P < 0.05$ small-volume corrected for family-wise error; images thresholded at $P < 0.005$, uncorrected for display].
To test our first prediction that decision bias is related to hippocampal activity, we used a general linear model to compare blood oxygen-level dependent (BOLD) activity during the Reward phase between items that led to later behavioral decision bias vs. those that did not, within individuals (i.e. \( S_1^+ > S_1^- \)) (Figure 3.2B). Activation in the posterior hippocampus during reward learning was greater for items that led to more decision bias (Figure 3.2C). A similar pattern was found across individuals: Activation in the hippocampus correlated with the proportion of biases in subsequent decisions (30, -6, -20; \( P < 0.05 \) SVC). As would be expected if decision bias is driven by the spread of value through associative memory processes, neural predictors of bias were selective to the Reward phase. Parallel analyses of BOLD activity during the Association and Decision phases found no areas of the brain within or across participants where activation correlated with decision bias.

To investigate whether hippocampal activity was related to explicit memory for the \( S_1-S_2 \) associations, after scanning we tested participants’ ability to correctly pair \( S_1 \) and \( S_2 \) items and asked about their choice strategy and awareness of task structure. Reflecting the automatic nature of the underlying associative memory processes (Turk-Browne et al., 2009; Wimmer and Shohamy, 2011), we found no evidence for explicit memory of the associations. Moreover, there was no relationship between measures of explicit \( S_1-S_2 \) memory and either decision bias or hippocampal activity (Figure 3.S3). While it is difficult to conclusively determine implicitness of cue pairings (De Houwer and Moors, 2007), these findings suggest that the role of the hippocampus in decision bias does not seem to be driven by explicit or strategic effects.
Figure 3.3. Visual reactivation predicts bias.

Reactivation of category-specific visual areas during the Reward phase predicts subsequent decision bias. 

A, Example participant region of interest masks (derived from the Association phase) for body, face, and scene S₁ stimuli. Masks were applied to S₂ presentations during the Reward phase. 

B, S₂ presentation elicits activation in visual regions responsive to associated S₁ stimuli when participants later exhibit decision bias ($t_{(22)} = 2.29$, $P < 0.05$).

To test our second prediction that decision bias is supported by reactivation of associations, we exploited the fact that the different categories of S₁ stimuli (faces, scenes, and body parts) elicit activation in distinct areas of visual cortex (Reddy and Kanwisher, 2006). In our design, these category-specific S₁ items were associated with S₂ items in the Association phase; in the Reward phase, however, only S₂ stimuli were presented. Thus, during reward learning, any selective activation in these visual cortical areas likely reflects associative reactivation, in memory, of these items. This allowed us to test whether biases in decisions, as measured behaviorally, were predicted by differential activation in category-specific areas during the Reward phase.
We analyzed associative reactivation during the Reward phase using participant’s responses to S₁ visual stimuli during the Association phase to define activation masks (Figure 3.3A). These participant-specific masks were then applied to Reward phase responses evoked by S₂ stimuli, relative to the alternative categories, resulting in a measure of reactivation for each participant for each association. We compared reactivation for associations that led to high vs. low decision bias. This analysis revealed that reactivation in visual regions during reward learning predicted later biases in decisions (Figure 3.3), such that across all categories there was greater neural reactivation for high- vs. low-bias decisions.

Reflecting the continual updating of memory representations, reactivation was significant in the first half of the Reward phase; it was present but weaker over the full Reward phase, as additional learning about the S₂ items took place. Importantly, reactivation was selective to the associated category-specific regions and was not general to all visual regions of interest.
Decision bias was related to functional connectivity between the hippocampus and the striatum during the Reward phase. A PPI analysis revealed correlated activity between the hippocampus and the striatum during trials that led to high vs. low decision bias. The same region in the striatum was also found to correlate with reward learning. This indicates that the hippocampus and striatum may form a functional circuit to support shifts in value.

Finally, we tested our third prediction that decision bias is related to functional interactions between the hippocampus and the striatum during the Reward phase. Specifically, we reasoned that an interaction between associative reactivation of $S_1$ stimuli via the hippocampus with reward learning in the striatum could provide a link between $S_1$ stimuli and reward.

We investigated functional connectivity during the Reward phase by conducting a psychophysiological interaction (PPI) analysis using right hippocampus as a seed and decision bias (high vs. low) as the behavioral (psychological) modulator. We found a
functional relationship between the hippocampus and the striatum which was
significantly greater for high-bias stimuli \( (z = 3.23 (6, 6, 12); P < 0.05 \) SVC; **Figure 3.4**).

We then investigated whether the correlation between the hippocampus and
striatum for high-bias decisions is mediated by reactivation in visual cortical regions. To
test this, we extracted trial-by-trial measures of evoked responses to high-bias stimuli and
performed a formal mediation analysis on the path between the hippocampus and the
striatum via the potential visual cortex mediator. Consistent with the PPI result, we
found that high decision bias was related to correlated BOLD activity in the hippocampus
and striatum \( (0.214 \pm 0.054, P < 0.001) \). Additionally, we found a trend for a mediating
effect of visual reactivation on this relationship \( (0.010 \pm 0.006, P < 0.08) \).

To test for the specificity of these connectivity results, we conducted a series of
control analyses. In a PPI analysis of the Association phase, we found no areas with
differential connectivity that related to decision bias, and in the Reward phase, we found
no significant activation outside of the striatum. Furthermore, in the mediation analysis, a
control region activated by Reward phase trials (in the temporal-parietal junction) did not
show any connectivity with regions of interest for high decision bias stimuli. Together,
these results indicate that decision bias depends not only on hippocampal activation
during reward learning, but additionally on functional connectivity between the
hippocampus and the striatum, putatively mediated by reactivation in visual cortical
regions.

These results indicate that reward can spread to bias the value of options that were
themselves never directly rewarded (Walther, 2002; Wimmer and Shohamy, 2011). This
provides insight into how people are biased by past experience to make new decisions.
between options that were never previously rewarded: Networks of associations in memory, formed across many different experiences, can result in the spread of value across associations.

The idea that memory and decision making are intertwined has deep roots in behavioral theories of both memory (Tolman, 1948; Eichenbaum and Cohen, 2001; Shohamy and Adcock, 2010) and decision making (Gilovich et al., 2002; Weber et al., 2007; Morewedge and Kahneman, 2010; Kahneman, 2011). Yet, there has been remarkably little evidence for an underlying mechanism. Understanding the mechanism by which value spreads among related memories in the brain may help explain why people sometimes develop seemingly ungrounded preferences for, or against, particular things, places, or people. Although we highlight how transfer of past experience can guide behavior in a changing environment, this same mechanism may also lead to seemingly “irrational” choices, consistent with social and cognitive theories regarding the role of associative memory in decision making heuristics (Gilovich et al., 2002; Kahneman, 2003; Weber et al., 2007; Morewedge and Kahneman, 2010; Kahneman, 2011).

The finding that the hippocampus supports the spread of value provides several new insights into the neural bases of both memory and decision making. First, our findings extend the role of the hippocampus beyond memory per se, demonstrating that the hippocampus contributes directly to value assignment and to decision making.

Second, although in humans the hippocampus is traditionally associated with explicit declarative memory (Eichenbaum and Cohen, 2001), our results indicate that transfer of value by the hippocampus is not driven by conscious awareness. Thus, our
results suggest that the hippocampus contributes to automatic assessment of value, perhaps performing a function similar to Bayesian inference about value (Tenenbaum et al., 2011).

Finally, while it is known that memory can support decisions by retrieving relevant information at the time of decisions (Preston et al., 2004; Johnson and Redish, 2007), our results demonstrate an alternative mechanism whereby the hippocampus dynamically modulates value representations during learning itself (Gluck and Myers, 1993; Eichenbaum, 2000; Shohamy and Wagner, 2008). This mechanism allows value to spread and bias decisions without effortful retrieval at the time of decision.

Understanding how associative memory biases decisions provides insight into critical open questions in decision making research. While reward learning models have been successfully applied to many aspects of behavior, these models cannot account for the full diversity of animal and human decision making (Daw et al., 2005; Hyman et al., 2006; Rangel et al., 2008). The uncovering of a neural mechanism by which associative memory biases decision making sheds light on how value generalizes across experience, with implications for both adaptive behaviors and maladaptive behaviors such as addiction.
Materials and Methods

Participants

Thirty-one fluent English speakers with normal or corrected-to-normal vision participated in the study. All subjects were free of neurological or psychiatric disorders and consented to participate. Informed consent was obtained in a manner approved by the Columbia University Institutional Review Board. Three subjects’ data were excluded because they did not show evidence of simple reward learning for the S2 stimuli, as measured in the post-learning Decision phase (less than 75% preference for all three reward-associated S2 stimuli). Behavioral and functional imaging data are presented from the remaining twenty-eight subjects (17 female; mean age, 23 years; range, 18-32). Subjects were paid $20 per hour for the approximately 2-hour duration of participation, plus one-fifth of the nominal rewards they earned in the experiment.

Procedure

The task was a newly developed variant of a “sensory preconditioning” paradigm (Brogden, 1939, 1947; Kimmel, 1977; Port et al., 1987). fMRI data were collected during three phases: Association, Reward, and Decision making (Figure 3.1). Additional behavioral data were collected before and after the task, as detailed below.

Pre-task stimulus ratings

Stimuli for each subject were selected based on a pre-task liking rating section. The stimulus selection procedure used subject’s liking ratings to select neutral stimuli that were closely matched in liking across categories to ensure that subjects did not have
strong pre-existing preferences for selected stimuli. Subjects rated 20 stimuli in each of the three picture categories (faces, scenes, and body parts) and 20 patterned circle stimuli using a scale anchored with “strong dislike” and “strong like”. Ratings data were collected during the high-resolution anatomical scan using an MRI-compatible trackball response unit; a subset of four subjects instead completed ratings on a laptop computer before entering the scanner. The selection algorithm picked two $S_1$ pictures from each of the face, scene, and body part categories and six $S_2$ circle stimuli for use in the six $S_1$-$S_2$ stimulus pairs in the experiment. (See supplementary text for descriptive statistics of pre-experiment stimulus ratings.)

**Association phase**

In the Association phase, subjects were incidentally exposed to sequences of pairs of stimuli (Figure 3.1A). On each trial, a picture ($S_1$; face, scene or body part) preceded a patterned circle ($S_2$). $S_1$ and $S_2$ stimuli were organized in pairs, with a particular $S_1$ stimulus always preceding a particular $S_2$ stimulus. Subjects were not informed of these pair relationships or the trial structure. To distract subjects from deliberately encoding the associations, they were given a cover task that instructed them to respond to occasional inverted target face, scene, or body pictures. (Targets were followed by a circle stimulus; neither the target picture nor circle were part of the critical stimulus pairs.) Subjects responded to target stimuli with a button press and to the $S_1$ and $S_2$ stimuli with an alternate button press. Each pair was presented 10 times in a pseudo-random order, intermixed with 18 target trials. To improve the implicit pairing of the stimuli, the inter-trial interval (ITI) was twice as long as the inter-stimulus interval (ISI) (4 s vs. 2 s;
Figure 3.S2) (following Walther, 2002; Wimmer and Shohamy, 2011). FMRI data were collected in two blocks of approximately 6 min duration.

**Reward phase**

In the Reward phase, subjects underwent a Pavlovian conditioning procedure in which S2 stimuli were predictive of either reward or neutral outcomes (Figure 3.1B). Only S2 stimuli were used as conditioned stimuli; no S1 stimuli appeared during this phase. Three S2 circle stimuli were paired with a reward (S2+) and the remaining three were paired with a neutral outcome (S2-). One S2 stimulus previously associated with each category of S1 (face, scene, and body part pictures) was paired with reward and one with neutral feedback, giving an S2+ and S2- stimulus for each category. Each S2 stimulus was presented on 16 reward learning trials in a pseudo-random order. For S2+ stimuli, the reward outcome appeared with a probability of 81% (13/16 trials); for S2- stimuli, the neutral outcome appeared on all trials.

To maintain continuity between phases, subject instructions were similar to those in the preceding Association phase. Subjects were instructed to respond to a target reward stimulus (a picture of a one-dollar bill) with a button press and the other stimuli (S2 stimuli and the neutral grey square stimulus) with an alternate button press. Subjects were instructed that when they correctly responded to the reward stimulus, the reward would be added to their earnings of which they would receive a percentage at the end of the study. Subjects were also informed that they might notice predictive associations between particular circle stimuli and reward or neutral outcomes.
A Reward phase trial consisted of the presentation of an $S_2$ stimulus, a fixation ISI, a reward or neutral outcome, and a variable ITI fixation of mean 2 s (range: 0.5-10.5 s). The duration and distribution of null events was optimized for estimation of rapid event-related fMRI responses as calculated using Optseq software (http://surfer.nmr.mgh.harvard.edu/optseq/). FMRI data were collected in two blocks of approximately 6.5 min duration.

*Decision phase*

In the Decision phase, we assessed subject’s preferences for the $S_2$ and $S_1$ stimuli. To assess reward learning, choices between $S_2$ stimuli included a reward-associated $S_2^+$ circle and a neutral-associated $S_2^-$ circle. To assess decision bias, choices were between $S_1^+$ and $S_1^-$ stimuli, i.e., between an $S_1$ stimulus from an Association phase $S_1$-$S_2$ pair where the incidentally associated $S_2$ was later rewarded ($S_1^+$) and an $S_1$ where the associated $S_2$ was not rewarded ($S_1^-$) (Figure 3.1C). Choices between $S_1$ stimuli were always within-category (e.g. between two face pictures). Subjects were instructed to choose the option that they thought was more likely to lead to winning $1. Subjects were informed that the outcome of their choice would not be presented immediately and that they would receive a percentage of their earnings at the end of the experiment.

On each trial, two stimuli were presented (randomly permuted on the left and right side). Subjects selected the left or right picture with a corresponding left or right button response during a 2.5 s response period. From the time the subject made a selection until the end of the choice presentation period, the selected option was framed in blue, followed by a jittered ITI with a mean of 2 s (range: 0.5-10.5 s). If no choice was
recorded during the choice period, the options remained on the screen until the end of the trial. Each critical choice between reward- and neutral-associated stimuli was presented four times, in pseudo-random order, yielding 24 trials. FMRI data were collected in one block of approximately 5 min duration.

*Post-task stimulus ratings*

After the completion of the Decision making phase and while still in the scanner, subjects were given a liking rating test similar to the ratings at start of the experiment, in order to test whether there were any changes in liking ratings. Subjects were presented with each of the $S_1$ and $S_2$ stimuli from the experiment and were instructed to indicate of how much they liked each stimulus.

*Explicit memory test*

Outside of the scanner, subjects completed a questionnaire on a laptop computer that probed memory for the $S_1$-$S_2$ pair relationships from the initial Association phase of the experiment. A single $S_1$ face, scene, or body picture was shown above two $S_2$ circle stimulus options: one that had been incidentally paired with the presented $S_1$ stimulus in the initial Association phase of the experiment and a lure that had been paired with a different $S_1$ stimulus. Subjects were instructed to select the circle ($S_2$) that seemed “related to” the picture ($S_1$).

Next, subjects completed a computerized questionnaire that instructed them to rate the likelihood that the $S_2$ and $S_1$ stimuli would be associated with winning money. Subjects then answered a series of written questions that assessed both memory for and
awareness of patterns of presentation during the Association phase and Reward phase, and choice strategies for the Decision making phase (see supplementary text). Finally, subjects were paid for their participation and given one-fifth of their winnings from the Reward and Decision making phases of the experiment.

Stimuli

All phases of the task were presented using the Matlab (Natick, Massachusetts) and the Psychophysics Toolbox (Brainard, 1997). Stimuli were projected onto a mirror above the subject’s eyes in the MRI system. The face stimuli were selected from the Stanford Face Database and the CVL Face database (Peter Peer, http://www.lrv.fri.uni-lj.si/facedb.html). Scene stimuli were selected from an internal database. Body part stimuli were selected via a search of publicly available images on the internet.

Imaging procedure

Whole-brain imaging was conducted on a 3.0T Phillips MRI system at Columbia University’s Program for Imaging and Cognitive Sciences, using a SENSE head coil. Head padding was used to minimize head motion; no subject’s motion exceeded 2 mm in any direction from one volume acquisition to the next. Structural images were collected using a high-resolution T1-weighted MPRAGE pulse sequence (1 X 1 X 1 mm voxel size). Functional images were collected using a gradient echo T2*-weighted echoplanar (EPI) sequence with blood oxygenation level-dependent (BOLD) contrast (TR = 2000 ms, TE = 20 ms, flip angle = 72, 2 X 2 X 3 mm voxel size; 38 contiguous axial slices).
For each functional scanning run, five discarded volumes were collected prior to the first trial to allow for magnetic field equilibration.

**Behavioral analysis**

We analyzed subjects’ responses during the Decision phase of the task and during the post-task tests (stimulus liking and explicit memory). We used subjects’ tendency to choose $S_2^+$ over $S_2^-$ stimuli as a measure of how well they learned the association between the $S_2$ stimuli and reward or neutral feedback in the Reward phase. Decision bias was estimated by measuring subjects’ tendency to choose $S_1^+$ over $S_1^-$ stimuli. A single preference score for each of the three reward-associated $S_2^+$ and the three associated $S_1^+$ stimuli was derived by averaging subject’s responses over the four presentations of each choice. Since any spreading of value to $S_1^+$ stimuli could be limited by the individual subject’s maximal preference for the $S_2^+$ stimuli, we also calculated a relative score of decision bias on a per-subject basis. This was computed by dividing $S_1^+$ choices by each subject’s average preference for the $S_2^+$ stimuli; if $S_1$ decision bias was greater than $S_2$ stimulus preference, the relative decision bias was set to 100%.

**Imaging analysis**

Preprocessing and data analysis was performed using AFNI (Cox, 1996) and Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were coregistered manually using AFNI. In SPM, images were realigned to correct for subject motion and then spatially normalized by estimating a warping to template space from each subject’s
anatomical image and applying the resulting transformation to the EPIs. Images were resampled to 2 mm cubic voxels, smoothed with an 8 mm FWHM Gaussian kernel, and filtered with a 128 s high-pass filter. SPM was used to estimate general linear models (GLMs) and psychophysiological interaction (PPI) analyses. Reactivation and mediation functional connectivity analyses were completed using AFNI and custom routines in Matlab.

For analysis of all phases of the experiment, fMRI model regressors were convolved with the canonical hemodynamic response function and entered into a general linear model (GLM) of each subject’s fMRI data. The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the GLM to account for residual effects of subject movement. Linear contrasts of the resulting SPMs were taken to a group-level (random-effects) analysis. We report results corrected for family-wise error (FWE) due to multiple comparisons (Friston et al., 1993); this approach assesses the strength of activations defined by an initial and arbitrary uncorrected threshold, which we take as $P < 0.005$ for all analyses. Accordingly, for display purposes, we render all SVC significant activations at this threshold. We conduct this correction at the peak-level within small volumes for which we had an a priori hypothesis or at the whole-brain cluster level. For regions of interest in the striatum and in the MTL (including both the hippocampus and parahippocampal cortex) we used anatomically defined masks derived from the AAL atlas (Tzourio-Mazoyer et al., 2002). Additionally, for the anterior hippocampus, a 6 mm diameter spherical region of interest was drawn at the coordinates reported previously in a different temporal association learning paradigm (28, -10, -22) (Staresina and Davachi, 2009). All voxel locations are reported in Montreal
Neurological Institute (MNI) coordinates; results are displayed overlaid on the average of all subjects’ normalized high-resolution structural images.

**Decision bias-related activation**

Our initial fMRI analysis focused on detecting activation directly related to decision bias during the Reward phase. The experimental design involved three different pairs of stimuli. Subjects often showed variability in decision bias across the face, scene, and body S₁ stimulus categories (Figure 3.2B and supplementary text). This variability allowed us to contrast the BOLD signal (hereafter referred to as “activation”) related to high versus low decision bias for each subject. This analysis included all subjects who exhibited variability in bias, excluding three subjects with 100% bias for all categories and two subjects with 0% bias for all categories. Behavioral decision bias, as a percent of Decision phase choices, was transformed within-subject into a rank measure, yielding a 1 to 3 low- to high-bias scale. The decision bias regressor in the GLM modulated predicted trial-evoked activation (6 s) during the Reward phase. A similar model was applied to the Association learning phase and the Decision phase data to test the specificity of Reward phase results.

**Reactivation of associations and decision bias**

Next, we explored within-subject correlates of decision bias that may be evident in Reward phase reactivation of paired S₁ stimuli. This analysis tested whether decision bias is predicted by reactivation of associations, which may allow the binding of reward and neutral outcomes to both the presented S₂ stimulus and reactivated S₁ stimulus. Our
paradigm utilized three categories of S₁ stimuli (faces, scenes, and body pictures), which have been found to activate distinct regions of the visual cortex: fusiform face area (FFA), parahippocampal place area (PPA), and extrastriate body area (EBA) (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Downing et al., 2001). This allowed us to test if activation during the Reward phase, where only S₂ stimuli were presented, also elicited activation of the category-specific S₁ stimuli incidentally associated with individual S₂ stimuli during the preceding Association phase.

To measure reactivation, we derived stimulus category (face, scene, and body part) masks for each subject from the Association phase, applied these masks to contrasts of S₂ stimuli in the Reward phase, and sorted the resulting values by subsequent decision bias. In detail, first, regions of interest were derived from activation to the stimuli presented in the Association learning phase. A GLM was estimated with separate regressors for S₁ face, scene, and body part stimuli as well as “upside-down” target trials (1.75 s). Contrasts were constructed to estimate responses to specific S₁ categories: [face - (scene + body)], [scene - (face + body)], [body - (face + scene)]. The resulting individual contrasts were masked to include only the top 1% of voxels for each contrast that also fell within a group mask (thresholded at P < 0.001, uncorrected) based on activation to category-specific stimuli in the Association phase. For face stimuli, we used a mask of the ventral visual cortex, as the group mask did not capture individual face-responsive regions in the ventral occipital cortex (Figures S4-S5). The resulting summed subject masks for face, scene, and body part stimuli resemble expected FFA, PPA, and EBA activation patterns, respectively, based on prior literature (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Downing et al., 2001) (Figures S4-S5).
Next, a Reward phase GLM was estimated similar to the model of the Association phase described above. Here, instead of modeling the presentation of $S_1$ face, scene, and body part stimuli, the GLM modeled the presentation of $S_2$ circle stimuli (2 s) that were incidentally paired with the face, scene, and body part pictures in the Association phase. Thus, the contrasts were between $S_2$ stimuli that differed only in their $S_1$ associations, e.g. $[S_2 \text{ face} - (S_2 \text{ scene} + S_2 \text{ body})]$. The category-specific masks from the Association phase were applied to the resulting contrasts of the Reward phase GLM. In the Reward phase, beta values in voxels falling within the mask were averaged to produce one value per $S_2$-associated category per subject. These values were sorted within-subject according to subsequent decision bias. Across-subject values for high, medium, and low bias were created from the sorted within-subject values. Statistical tests were performed on the resulting data for high vs. low bias. The majority of subjects had two equivalent high- or low-bias pairs, which were averaged together; the medium bias condition only existed within a subset of subjects ($n = 8$) and was not included in subsequent analyses due to lack of statistical power.

The Association phase masks were also used to compute mean levels of activation in the Association phase contrasts themselves. These values were used to establish the specificity of Reward phase reactivation in two ways. First, they were verified to not differ according to later decision bias. Second, they were used to normalize Reward phase reactivation estimates by controlling for baseline levels of responding to the actual picture stimuli in the Association phase. It could be argued that the use of the Association phase for voxel selection induces a confound, given the inclusion of both the $S_1$ and associated $S_2$ stimulus on each trial. However, any bias induced by the Association phase
will be minimized as the Association phase activation was itself used to normalize the Reward phase reactivation estimates. Additionally, when a similar reactivation procedure was used with masks derived from the Association phase S2 presentation, Reward phase activation did not correlate with decision bias.

**Reward phase parametric learning signals**

We additionally tested for signals related to reward learning. FMRI data in the Reward phase were analyzed using event onsets that modeled the whole trial (6 s) as well as the onsets for the S2 stimulus (0.5 s) and the reward outcome (0.5 s).

To model reward learning, we utilized a standard temporal difference reinforcement learning model to correlate BOLD activity with trial-by-trial estimates of stimulus value and reward prediction error, the difference between stimulus value and received reward (O'Doherty et al., 2004; Daw et al., 2006; Hare et al., 2008). The Reward phase associations between S2 stimuli and outcomes were Pavlovian, with reward appearance independent of behavioral responses, and thus we could not fit the parameters of a reinforcement learning model to behavior. Instead, the reinforcement algorithm learning rate $\alpha$ was set at 0.25 such that stimulus values for S2 stimuli, initialized at zero, reached asymptote for S2+ stimuli (approximately 0.7) by the end of the phase. (Results are robust to different learning rate settings.)

To model association learning signals, we adapted a temporal difference model to correlate BOLD activity with trial-by-trial estimates of stimulus “surprise” due to the unpredicted presentation of lone S2 stimuli. During the preceding Association phase of the experiment, specific S1-S2 temporal associations were incidentally experienced, but in
the Reward phase, these associations are violated by the presentation of the lone $S_2$
stimulus. The unexpected omission of the $S_1$ and the unexpected lone presentation of the
$S_2$ in the Reward phase may lead to a “stimulus prediction error”, a response previously
attributed to the hippocampus and midbrain dopamine system (Lisman and Grace, 2005;
Kumaran and Maguire, 2006, 2007). Prediction errors are highest at the beginning of
reward learning when expectations of joint $S_1$–$S_2$ presentation are the highest. As the
Reward phase proceeds, expectations would be updated via the prediction error signal,
leading to decreasing surprise. The learning rate for this model was set to $\alpha = 0.25,$
similar to the reward learning rate. The prediction error signal thus captures stimulus-
specific activation that exponentially decreases with repeated $S_2$ presentations during
Reward phase trials.

*Functional connectivity analysis*

Two methods were used to explore functional connectivity related to decision
bias. First, we performed a psychophysiological-interaction (PPI) analysis (Friston et al.,
1997). Second, to test for mediation of decision bias-related connectivity, we extracted
per-trial estimates of evoked activation and conducted regression analyses (Adcock et al.,
2006). These analyses focused on the first half of the reward phase, where we found
evidence for significant reactivation.

In the PPI analysis, a seed region in the hippocampus was defined in the GLM
contrast of high versus low decision bias during Reward learning (26, -34, -12, 6 mm
region of interest; **Figure 3.2C**). The PPI analysis was estimated to test for increases in
functional correlation between the hippocampus and other brain regions (the physiological variable) during the Reward learning phase for high versus low decision bias stimuli (the psychological variable). The timecourse of activation from the hippocampus was extracted and deconvolved. This timecourse was multiplied by the decision bias indicator and then convolved with the hemodynamic response function (HRF). The model included the hippocampus timecourse by decision bias regressor, the decision bias regressor and the unmodulated hippocampus timecourse regressor (Friston et al., 1997). We explored interactions in the whole brain and in striatal regions of interest defined by 6mm radius spheres around peak striatal clusters correlated with reward prediction error (6, 6, 12) and value (18, 14, 4) in the reward learning GLM (Figure S3.3). As a control, we analyzed data from the Association phase in a similar manner.

Finally, we tested whether visual regions, defined per-subject, mediated decision bias-related connectivity between the hippocampus and striatum. While the visual cortex pathway lacks direct connectivity with the other regions, functional correlations could arise via intermediate pathways. This analysis was conducted on trial-by-trial extracted activation values which enabled the use of subject-defined masks as regions of interest. Stimulus-evoked activation to S_2 stimuli for each trial was estimated by extracting normalized activity from 4-6 s after trial onset, minus a baseline response taken from -2 to 0 s before trial onset. The response timepoint selection was visually verified to overlap with peak trial-evoked responses independent of condition. Activation values were sorted into two bins within each subject: S_2 stimuli for which the associated S_1 stimulus exhibited subsequent decision bias versus those that did not. This method is conceptually similar to an approach that estimates separate regressors for each trial in a GLM and
combines the resulting estimates to compute between-region correlations (Rissman et al., 2004).

As in the PPI analysis, activation values in the hippocampus were extracted from the region that exhibited activity predictive of decision bias; activation values in the striatum were extracted from the ROI in the right caudate correlated with reward prediction error. For the visual cortex, we extracted activity from subject-specific ROI masks derived from face, scene, and body responses, as used in the reactivation analysis. As a control, we extracted data from an ROI in the temporal-parietal junction (62, -42, 22) that exhibited a significant response to Reward trials overall but which would not be expected to contribute to decision bias.

To test whether visual regions mediate connectivity between the hippocampus and striatum, we conducted a formal mediation analysis (Wager et al., 2009). A multi-level mediation was estimated using a custom Matlab toolbox (Tor Wager, http://wagerlab.colorado.edu/tools). We estimated per-subject path coefficients for the hippocampus to striatum (Path c), hippocampus to visual cortex (Path a), and visual cortex to striatum (Path b). The mediation tests whether the path via the visual cortex accounts for significant covariance in the hippocampus-striatum path. At the second-level, the analysis uses bias-corrected bootstrap significance testing to derive a sensitive measure of path and mediation effects (randomly sampling 100,000 observations within each path at the subject level). This analysis provides path coefficients, standard errors, and two-tailed, uncorrected p-values for each path and the mediation effect.

Association and Decision phases
Where appropriate, parallel GLM and functional connectivity analyses were conducted on data from the Association and Decision phases of the experiment comparing activation that led to high vs. low decision bias, as noted above.
Supplementary Text

Pre-experiment liking ratings

Stimuli were selected based on individual ratings of 20 stimuli in each category to match median liking across stimulus types. On the liking scale, transformed into 0-100% values, the mean liking for selected S₁ stimuli and S₂ stimuli was almost exactly matched (56.14% ± 1.73 and 56.11% ± 1.66, respectively (mean ± SEM)). Within S₁-S₂ pairings, stimuli were also closely matched in liking: the absolute difference in liking between stimuli in a pair across subjects was near zero (mean participant rating deviation, 1.78%; maximum, 6.75%). Residual initial liking rating differences between S₁ stimuli did not predict subsequent preferences during the Decision phase. Liking ratings for subsequently chosen high-bias (S₁+ versus S₁-) stimuli were no different than zero (-0.05% ± 0.61). Liking ratings were numerically lower for low-bias stimuli (-1.66% ± 0.92), a pre-experiment difference in liking ratings that would, if anything, work against a subsequent decision bias.

Association learning phase

During the Association learning phase, performance on the target-detection task was near ceiling (2.78% ± 1.01 incorrect responses; 4.70% ± 1.22 missed responses). Reaction times to S₁ or S₂ stimuli did not differ according to later pair memory or decision bias.
**Reward learning phase**

During the Reward learning phase, responses to the reward were rarely missed or in error (misses, 1.44% ± 0.77, errors, 0.92% ± 0.35 of trials). Reaction times to the $S_2$ stimuli were 778.3 ms ± 29.0. Reaction times to $S_2^+$ and $S_2^-$ stimuli did not differ; both showed a decrease over time, but this change only reached significance for $S_2^+$ stimuli (from 796.7 ms ± 30.2 to 759.9 ms ± 30.6; $t_{(27)} = 2.44$, $P < 0.05$). Reaction times to $S_2$ stimuli did not differ according to later decision bias.

**Decision making phase**

Reward-associated $S_2$ stimulus ($S_2^+$) preference overall was 79.4% ± 3.9. Broken down by associated $S_1$ stimulus picture type, mean preference values were 79.5% ± 6.7 for face-associated $S_2$ stimuli, 81.3% ± 5.5 for scene-associated $S_2$ stimuli, and 76.8% ± 7.0 for body-associated $S_2$ stimuli.

Decision bias overall, as measured by $S_1^+$ versus $S_1^-$ preference, was 53.8% ± 5.4 (computed relative to $S_2$ preference). Broken down by picture type, mean preference values were 54.2% ± 9.4 for face $S_1$ stimuli (n.s.), 42.0% ± 8.9 for scene $S_1$ stimuli (n.s.), and 66.4% ± 8.2 for body $S_1$ stimuli ($t_{(26)} = 1.99$, $P < 0.10$). Within-participant, ranking the three stimulus pairs of different categories by decision bias illustrates variability that was subsequently used to probe neural correlates of bias (86.3% high, 50.6% medium, 20.8% low decision bias; Figure 3.2B). Each category type only allowed choices versus one alternative stimulus in the Decision phase and thus participants were often at 100% or 0% bias for a given category. Among ranked decision bias values, for the high stimulus 19/28 (68%) of participants exhibited 100% bias for at least one pairing and
23/28 (82%) of participants show greater than 50% preference for at least one pairing. Conversely, for at least one low stimulus, 21/28 (75%) of participants exhibited 0% bias.

Decision making phase reaction times were 1228.8 ms ± 46.2 for S\textsubscript{1} choices and 1164.3 ms ± 4.7 for S\textsubscript{2} choices. Reaction times for S\textsubscript{1} and S\textsubscript{2} choices did not significantly differ. Within-participants, reaction times did not differ between high versus low decision bias S\textsubscript{1} choices.

In the Reward learning and Decision making phases of the experiment, participants earned an average of $14.00 ± 1.41.

*Post-experiment liking ratings*

After the Decision making phase, participants rated their current liking for the S\textsubscript{1} and S\textsubscript{2} stimuli. (Liking data for one participant is missing due to an error in data recording.) While this measure may not be fully independent of Decision phase preferences (as choices themselves may influence liking), results from this phase nevertheless support the preferences exhibited in the Decision phase. Liking for S\textsubscript{2} stimuli showed a strong effect of reward associations in the Reward phase (change in S\textsubscript{2}+ minus change in S\textsubscript{2}-, 15.17% ± 3.63; t\textsubscript{26} = 4.18, P < 0.001). Liking overall did not increase over the experiment for S\textsubscript{1}+ versus S\textsubscript{1}- stimuli (1.26% ± 1.65, n.s.). However, liking for S\textsubscript{1} stimuli when sorted by decision bias exhibited a significant difference for high versus low decision bias stimuli (change in S\textsubscript{1}+ minus change in S\textsubscript{1}-; high = 8.95% ± 3.46; low = -3.26% ± 2.33; t\textsubscript{22} = 5.08, P < 0.01; Figure S3.3A).

A final measure further supports the Decision phase choice and liking ratings. When outside of the scanner, participants were asked to rate, for each S\textsubscript{1} and S\textsubscript{2} stimulus,
how likely they thought it would be to be associated with winning money. (Participants 
rated stimuli on a scale similar to that used for liking ratings, but labeled with 0\% reward 
and 100\% reward anchors.) Reward likelihood ratings for S\_2 stimuli reflected reward 
associations in the Reward phase. Reward likelihood ratings for S\_1 stimuli when sorted 
by decision bias exhibited a significant difference for high versus low decision bias 
stimuli (S\_1+ minus S\_1-; high = 17.15\% ± 4.85; low = -11.59\% ± 5.53; t\(_{22}\) = 4.46, P < 
0.001), consistent with the results from the liking rating questionnaire.

*Testing explicit awareness of associations*

After leaving the MRI scanner, participants completed a matching probe and a 
written questionnaire to determine memory and post-experiment awareness for the 
incidental S\_1-S\_2 associations from the Association phase of the experiment (see Methods 
for details). (Data from one participant is missing due to an error in data recording.) Pair 
memory accuracy for S\_1-S\_2 associations was not different than chance (53.7\% ± 3.25 
correct; only one participant exceeded 4/6 correct). Pair memory averaged across picture 
categories did not correlate with mean decision bias, nor did pair memory within-
category correlate with decision bias within-category. When pair memory accuracy was 
sorted by high versus low bias, the difference was not significant (high 59.38\% ± 7.3 vs. 
low 45.7\% ± 7.5; t\(_{22}\) = 1.19, n = 23, P > 0.2; Figure S3.3B).

As noted in the main text, exploratory analyses of brain activation related to 
association memory in the Association and Reward phases revealed no activation in 
regions of interest to correctly-associated versus incorrectly-associated stimulus pairs, 
even at a liberal uncorrected threshold of P < 0.01.
On the final paper questionnaire, we asked participants if they noticed any regularities or pairings in the first (Association) part of the experiment. (For text of questions, see section at end of supplement.) No participants exhibited evidence of awareness of the pairings. Finally, in an additional question, no participants exhibited knowledge of the hypothesis of the study. Combined with participants’ pair memory accuracy, these data indicate that participants’ Decision phase choices and activation in the hippocampus were not driven by explicit awareness of stimulus associations.

However, it is possible that participants’ awareness of associations following the Association phase was higher than that exhibited after the experiment; such awareness could influence participants’ learning and decision making. While we cannot probe memory in the middle of the experiment without drawing attention to the manipulation of interest, we conducted a separate additional behavioral study to address this question. In this experiment, instead of probing memory for pairings at the end of the experiment after the Reward and Decision phases, pair memory was probed immediately following the Association phase. Significant memory performance and pair association awareness at this point could indicate that participants in the fMRI study also had access to stronger association memories during learning and decision making. Procedures for the pre-experiment stimulus rating, Association phase, and pair memory probe were otherwise identical to those in the main fMRI study. Immediately after the Association phase, participants (n = 19, 11 female, age range 18-26 years) completed the pair memory probe that instructed participants to match an $S_1$ stimulus at the top of the screen with one of two $S_2$ stimuli below.
Replicating the finding from the main fMRI experiment, as well as a previous behavioral study (Wimmer and Shohamy, 2011), pair memory performance was not different from chance (56.1% ± 5.4; $t_{(18)} = 1.13, P > 0.20$). Further, in an extended written questionnaire, no participants reported noticing the regular $S_1$-$S_2$ pair presentations. When informed about the pairs and then asked if they noticed the pair presentations, the majority of participants responded in the negative (11 responded “no”, 4 responded “a little”, and 4 responded “yes”). Memory performance in the subgroup that reported having noticed the pair presentations was lower than in the full group (41.5%, $n = 4$). After being informed about the Association phase pairs, participants were asked which type of picture (circle, face, scene, or body) was always present in a pair. 75% of participants (of 16 who selected an answer) correctly responded that the circle was always present. We then asked participants to identify whether the circle came first or second in the sequence: Only half of those who had responded correctly that the circle was always present answered correctly that the circle came second. Participants’ inability to recognize the temporal position of the common stimulus, immediately subsequent to an Association phase with 60 repetitions of pairings, indicates a low level of awareness for the association structure.

The lack of significant pair memory supports the hypothesis that the kind of association learning engaged during the task does not tend to lead to explicit, declarative knowledge of associations. Thus, with respect to mechanisms underlying decision bias, these null memory findings suggest that strategic inferential reasoning about $S_1$-$S_2$-reward associations, operating during the Reward phase or the Decision making phase, is unlikely to account for shifts in $S_1$ preferences.
Nevertheless, it is possible that the memory probes we employed did not reveal awareness of the pair associations underlying the task. Some studies have demonstrated that awareness during learning may not be reflected in memory measures (c.f. Gawronski and Walther, 2012). Thus, determining the degree of implicitness underlying decision bias is an important area for future study.

*Decision bias predictors*

During the Reward learning phase, outside of the hippocampus, a whole-brain FWE-corrected analysis of the decision bias contrast did not show any additional regions predictive of bias. In the striatum, we did not find any regions predictive of bias, even at a liberal uncorrected threshold of $P < 0.01$.

*Reactivation predictors of decision bias*

In the first half of the Reward learning phase we found significant reactivation of associations for stimuli in pairs that later showed high vs. low decision bias, as described in the main text (*Figure 3.3*). We found that over all categories, reactivation was greater for high vs. low decision bias in 74% of participants ($t(22) = 2.55, P < 0.05$). The reactivation effect was present numerically across the entire Reward learning phase (high = $0.121 \pm 0.039$, low = $-0.020 \pm 0.081$; $t(22) = 1.13, P < 0.20$). When limiting our analyses to the second half of the Reward phase a difference was not found (high = $-0.006 \pm 0.041$, low = $0.014 \pm 0.044$; $t(22) = 0.33, n.s.$).

In a follow-up analysis we examined activation to each stimulus category separately (faces, scenes and body parts), comparing individual associations that led to
later high vs. low decision bias. As this analysis is restricted to a separate test for each category, activation values for high vs. low decision bias are necessarily compared between-participants and include a smaller group of participants, both of which lead to weaker statistical power to detect an effect. For face-associated S2 stimuli, the reactivation measure was greater for high vs. low decision bias trials at a trend level (high \(0.238 \pm 0.110, \text{low } -0.034 \pm 0.075; t_{(19)} = 2.08, P < 0.06; n_{\text{high}} = 10, n_{\text{low}} = 11\)).

Activation for body-associated S2 stimuli was numerically higher for high vs. low decision bias (high \(0.056 \pm 0.043, \text{low } -0.060 \pm 0.013; t_{(17)} = 1.08, P > 0.10; n_{\text{high}} = 13, n_{\text{low}} = 6\)), while in the scene-associated S2 stimuli we detected no difference (high \(0.003 \pm 0.060, \text{low } 0.009 \pm 0.065, n.s.; n_{\text{high}} = 7, n_{\text{low}} = 14\)).

As a control, we tested whether reactivation was due to generally higher levels of activity across visual regions of interest rather than in specific visual regions reflecting S2 pairing. Mean activation across the three sets of participant-specific masks did not differ for high vs. low decision bias Reward phase trials (\(P > 0.45\)). Also, we verified that the reactivation result was found if category-specific region of interest activation was normalized by activation across all visual regions of interest. This measure exhibited much higher variance (for example, activation to actual visual stimuli in the Association phase was only marginally significant), but nevertheless, the reactivation effect was replicated (69% of participants exhibited greater reactivation for high versus low decision bias).

Finally, as noted in the main text, as a control we also tested whether activation in the Association learning phase predicted decision bias. While we found activation overall during the Association learning phase in the hippocampus and surrounding medial
temporal lobe (Figure 3.S2, masked for the MTL), we did not find significant correlates of decision bias in the MTL. Similarly, during the Association phase, activation to the presentation of S1 face, scene, and body stimuli did not differ by later bias ($t_{(22)} = -0.78$, n.s.). A parallel analysis of reactivation in the Decision phase is not possible as choices present actual face, scene, and body part stimuli.

**Reward learning activation**

In the Reward phase reward learning GLM, we examined brain activation related to learning associations between S2 stimuli and reward and neutral outcomes. The reward learning signals we examined were derived from a simple temporal difference model of learning (e.g., Sutton and Barto, 1998; O'Doherty et al., 2004). At S2 stimulus presentation we tested correlates of learned stimulus value, the reward prediction error elicited by stimulus presentation; at reward presentation we tested correlates of reward prediction error, focusing on the striatum (O'Doherty et al., 2004).

Activation in the bilateral striatum significantly correlated with the predicted value of presented stimuli (Figure S3.3A). At reward presentation, a region of the right caudate correlated with reward prediction error (Figure 3.3B; 90 voxels at $P < 0.005$, uncorrected). While this activation did not survive small-volume correction in the striatal region of interest ($P < 0.12$), we focus on this region in later analyses given its proximity to the striatal region correlated with value. Additionally, a region of the right ventral putamen significantly correlated with reward prediction error (Figure S3.3B).
In an exploratory analysis, we also found that activation in the posterior hippocampus correlated with reward prediction error (20, -28, -6; z = 3.86, p < 0.05 SVC, correcting for the additional ROI analysis).

**Reward phase association learning activation**

In this analysis, we found a cluster of activation in the ventral tegmental area / substantia nigra pars compacta (VTA/SNc) that significantly correlated with the stimulus prediction error regressor (z = 3.40 (4, -18, -12)). This cluster overlaps with a midbrain activation reported in a previous study of learning in a related paradigm (Shohamy and Wagner, 2008). Clusters of activation in the left and right hippocampus were significant at P < 0.001 uncorrected but did not survive small-volume correction. This result was specific to the Reward phase and was not due to generic decreases in activity across the phase.

**Functional connectivity analysis**

Results of the Reward phase PPI analysis of hippocampal-striatal functional connectivity for high- vs. low-bias stimuli are reported in the main text and Figure 3.4. A control analysis of hippocampal connectivity during the Association phase revealed no interaction between decision bias and hippocampal connectivity anywhere in the brain, even at a liberal uncorrected threshold of P < 0.01.

A summary of the Reward phase mediation result is reported in the main text. This analysis tested whether a pathway from hippocampus to striatum via participant-specific visual cortical regions mediated hippocampal-striatal connectivity. We report
mediation results from the full group of participants with high-bias stimuli, for the first half of the Reward phase. While this analysis is different from a PPI analysis and uses trial-by-trial estimates of evoked responses, we found that hippocampal-striatal connectivity replicated that found using PPI. More complete statistics on the paths between regions are as follows: Path a) hippocampus to visual cortex (0.426 ± 0.086, P < 0.001); Path b) visual cortex to caudate (0.050 ± 0.015, P < 0.001); Path c) hippocampus to caudate (0.214 ± 0.054, P < 0.001). For low-bias stimuli, functional connectivity between the hippocampus and visual cortex was not significant and we found no evidence for mediation.

To verify the specificity of the functional connectivity result between the hippocampus and the caudate in predicting decision bias in the Reward phase, we conducted several control analyses, as mentioned in the main text. First, a control analysis tested connectivity between regions of interest and a task-activated region of the temporal-parietal junction during the Reward phase that was not expected to show differential connectivity in the temporal-parietal junction. This analysis revealed that connectivity between the hippocampus and the control ROI did not differ between trials that led to high versus low decision bias.

As another control, we analyzed Association phase connectivity between regions of interest for high- and low-bias stimuli. We found no significant connectivity during trials that led to high bias for either the hippocampus to visual cortex pathway or the pathway between visual cortex and striatum. These results confirm the selective role of functional connectivity during the Reward phase between the hippocampus, visual cortex, and caudate in predicting subsequent decision bias.
Figure S3.1. Liking ratings and association memory.

Liking ratings and association memory. **A**, Change in liking ratings over the experiment plotted by level of decision bias among three stimulus categories. **B**, Post-experiment association memory accuracy for pairs (S₁-S₂ associations) plotted by level of decision bias.
Figure S3.2. Association phase activation in the hippocampus.

Association phase activation in the hippocampus. Right hippocampus, \( z = 5.67 \) (22, -28, -6); left hippocampus, \( z = 5.32 \) (-22, -26, -6). (All fMRI results \( P < 0.05 \) SVC for FWE unless otherwise noted and masked for regions of interest; images thresholded at \( P < 0.005 \), uncorrected for display.)
Figure S3.3. Reward learning-related activation.

Reward learning-related activation during the Reward phase. A, Activation in the striatum correlated with predicted value, left. Right striatum peak: caudate-putamen, $z = 4.31$ (18, 12, 4); left striatum peak: putamen, $z = 3.91$ (-18, 0, 16). B, Activation in the dorsal caudate correlated with reward prediction errors, right ($z = 3.50$, $P < 0.005$ unc., 90 voxels). Right ventral putamen, $z = 3.86$ (26, 6, -8).
Figure S3.4. Association phase group activation to visual stimuli.

Activation in the Association phase to face, scene, and body S1 stimuli (P < 0.001, uncorrected).
Figure S3.5. Association phase individual activation to visual stimuli.

Sum of participants’ category-selective masks derived from the S1 presentation in the Association phase and used for the Reward phase reactivation analysis. Individual masks represent the peak 1% of voxels falling within group Association phase-derived mask (see Methods). Color represents number of subjects with overlapping masks at each voxel (of n = 23 subjects); maximum n = 8, yellow; minimum n = 1, magenta.
Chapter 4: Episodic memory modulation by reward learning

Chapter 4 is from an in preparation manuscript: Wimmer, G.E., and Shohamy, D.
In Chapters 2 and 3 I demonstrated that relational representations impact reward learning and value-based decision making, via hippocampal-striatal connectivity. The studies in Chapters 4 and 5 address the complementary question of how long term episodic memory is impacted by ongoing reward learning.

Abstract

Memory is central to adaptive behavior. Extensive research has demonstrated that the cognitive and neural systems for remembering episodes are distinct from the ability to associate stimuli or actions with reward. However, in everyday life these two types of learning often co-occur. Currently, it is unknown whether and how memory for events is affected by ongoing learning of stimulus-reward associations. We thus sought to examine how these two processes interact – that is, how episodic memory is influenced by incremental learning of stimulus-reward associations. If these two types of learning are indeed distinct, then during a rewarded learning task, feedback would only influence learning, with no impact on memory for events. Alternatively, trial-by-trial choices and reward feedback may also influence episodic memory, indicating a functional interaction between these two processes. We developed a task where participants made choices between two options, each associated with drifting reward probabilities, and received feedback (a monetary reward or nothing). During learning, pictures of objects were presented as part of the choice options (Study 1) or as part of the feedback (Study 2). These pictures were not related to the reward task, and each object appeared only once during the experiment. Subsequently, participants were given a surprise memory probe to
measure episodic memory for the incidentally presented pictures. To test how reward learning affects long-term memory formation, the probe was administered immediately after learning or a day later. Although unrelated to prediction or receipt of reward, episodic memory for the pictures was above chance. Moreover, memory was influenced by trial-by-trial experiences. Study 1 revealed that memory for pictures presented at choice was enhanced by reward, but only when the memory test was immediate: surprisingly, we found that after a day reward impaired memory. In Study 2, we found that for pictures presented during feedback, reward enhanced memory, and this was stronger after a delay. These findings demonstrate that ongoing learning about reward can influence memory formation, suggesting novel interactions between episodic memory and reward learning.
Introduction

In everyday life, experiences of events and of reward outcomes are often intermixed and interacting. However, memory for events and rewards have traditionally been studied separately. Understanding the intersection of these types of learning and memory is important, as this may illuminate factors influencing the selectivity of memory. Further, memories formed during reward experiences may be especially relevant for future decision making.

Research on the cognitive and neural systems supporting episodic memory has focused on understanding memory for neutral events. These studies have revealed a critical role for the hippocampus and medial temporal lobe (Davachi, 2006). Research on the systems supporting reward learning and value-based decision making, on the other hand, has focused on how humans and animals learn from and utilize reward feedback. These studies have revealed a critical role for dopamine and the striatum (Rangel et al., 2008). Here, stimulus-reward associations are established via well-described reward learning signals conveyed by the phasic firing of midbrain dopamine neurons (Schultz, 2006). However, it is largely unknown whether and how episodic memory is influenced by learning from reward. For example, it is unknown if episodic memory is modulated by the same learning signal conveyed by dopamine neurons. Additionally, it is unknown if reward similarly modulates memory for experiences preceding feedback vs. experiences that coincide with feedback.

Prior studies have found memory enhancement by single-trial reward feedback or reward motivation (Wittmann et al., 2005; Adcock et al., 2006; Bialleck et al., 2011; Mather and Schoeke, 2011). These studies suggest that reward may influence both
ongoing incremental learning of stimulus-reward associations as well as episodic memory formation. Potentially, reward may influence memory by projections from midbrain dopamine neurons to the hippocampus (Shohamy and Adcock, 2010). Midbrain dopamine neurons projecting to the striatum update value associations via the difference between expected and received reward (reward prediction error) (Schultz, 2006), and this signal may also influence the encoding of episodic experiences in the hippocampus. Dopamine in the hippocampus has been shown to enhance cellular measures of memory over time (Huang and Kandel, 1995; Bethus et al., 2010). The possible influence of dopamine on episodic memory may be particularly evident after a delay, allowing for memory consolidation.

Similarly, the presentation of choice options with high predicted values increases dopamine neuron firing (Morris et al., 2006; Roesch et al., 2007), a signal which may also increase memory. In a choice situation, it is also possible that memory will be enhanced for experiences associated with choices. In addition to the influence of value, a beneficial effect of choice on memory may arise from effects of attention and self-guided exploration (Krajbich et al., 2010; Voss et al., 2011).

Research on memory and emotion also offers an important perspective on the modulation of memory by reward. Emotion, in particular arousal, has been shown in many cases to enhance memory (McGaugh, 2004). During ongoing reward learning, more surprising reward or miss feedback may elicit an increase in arousal, which may modulate memory. Prior research has found that arousing events can both increase (Anderson et al., 2006) and decrease (Mather, 2007) memory for nearby neutral events.
When arousing events are directly relevant for adjusting ongoing behavior, it is unknown whether arousal or reward (valence) will predominantly affect memory formation.

Alternatively, in an ongoing learning context, episodic memory may be unaffected by reward. In research on the function of multiple memory systems, a reward or “habit” learning system and an episodic memory system are often found to be independent or to interact competitively (Poldrack and Packard, 2003). Thus, incremental reward learning may be independent from memory formation. Further, memory may be unaffected by reward if feedback is allocated exclusively to updating the value of the choices, performing a function as in reinforcement learning models (Sutton and Barto, 1998).

To explore whether and how episodic memory is influenced by choices and reward feedback, we developed a novel reward learning task. Participants made repeated choices between two options and received trial-by-trial feedback. The reward probability associated with each option shifted throughout the game to encourage continuous learning. During each trial, incidental pictures of objects appeared within choice options or within feedback. Memory was assessed during a surprise memory probe.

In Study 1, incidental pictures were presented during choice; choice options were distinguished by squares with different color shading (green or blue) and objects were presented within them (Figure 4.1A). In Study 2, incidental pictures were instead presented as part of the reward feedback; reward or miss feedback was distinguished by different color shading (green or blue) (Figure 4.1B). Memory was tested immediately (Study 1A and 2A) or the following day (Study 1B and 2B) in separate groups of participants (Figure 4.1C). Our experimental design thus allows us to determine whether
episodic memory for events at choice or outcome are influenced by reward, and whether such effects vary as a function of delay.
Figure 4.1. Reward learning and memory task.

A, Reward learning task with choice pictures used in Study 1. B, Reward learning task with feedback pictures used in a separate group of participants in Study 2. C, Surprise subsequent memory test from Study 2. In separate groups, memory was probed immediately or after a day.
Methods

Participants. A total of eighty-nine participants completed Studies 1 and 2: Study 1A (Choice, immediate) included 20 subjects (12 female; mean age, 20 years, range, 18-25); Study 1B (Choice, delay) included 29 subjects (19 female; mean age, 21.1 years, range, 18-27) after excluding two with low corrected memory performance (less than 5%) and one who missed more than 20% of choices in the reward learning task; Study 2A (Feedback, immediate) included 20 subjects (10 female; mean age, 23.4 years, range, 18-34) after excluding three subjects with low memory performance and Study 2B (Feedback, delay) included 20 subjects (15 female; mean age, 21.2 years, range, 18-34) after excluding one with low memory performance. Subjects were paid $12 per hour for the approximately 1.5-hour duration of participation plus one-fifth of the nominal rewards they earned in the experimental task. Informed consent was obtained in a manner approved by the Columbia University Institutional Review Board.

Study 1: Memory for choice events. The experiments consisted of two sessions, either performed on the same day (Study 1A) or performed on consecutive days (Study 1B). In session 1, subjects completed a reward learning task that included incidental exposure to trial-unique picture stimuli during choice. In session 2, subjects completed a surprise test of memory for pictures that were seen during the reward learning task.

The reward learning task (Figure 4.1A) is a variant of a “two-armed bandit” task (e.g. Schönberg et al., 2007). Choice options were distinguished by a colored overlay (blue or green; Figure 4.1A). Subjects were instructed that each option was associated with a different probability of reward, that these probabilities could change slowly, and
that their goal was to attempt to find the most rewarding option at a given time in order to earn the most money. They were also instructed that rewards were tied to the color and not the screen position of the color. Subjects completed a short practice version to familiarize them with the task.

On each trial, subjects had 2 s to choose between the two options. The left-right location of the options was permuted randomly from trial to trial. So as to maintain equal visual saliency of both options and the included pictures, the selected choice was not highlighted. Options remained on the screen for 2.5 s. After a 1 s inter-stimulus-interval, subjects received binary reward feedback for 1.75 s. A $0.25 “win” was represented by an image of a quarter dollar and a $0.00 “miss” was represented by a phase-scrambled image of a quarter dollar (Figure 4.1A). If no choice was recorded during the choice period, no reward feedback was displayed and "Too late!" was presented on the screen until the end of the trial. Trials were separated by variable duration inter-trial interval (ITI) fixation null events (mean 3.25 s, range 1-10 s). To signal trial onset, the ITI fixation cross changed to black 1 s preceding the next trial. The experiment consisted of 80 choices.

Each of the options was associated with a different probability of monetary reward. Throughout the experiment, the reward probabilities diffused gradually to encourage continual learning. The reward probability for each option changed over time, diffusing between 20% and 80% according to Gaussian random walks with reflecting boundary conditions.

To investigate the influence of reward learning on episodic memory, incidental trial-unique object pictures were included in each option. Subjects were informed of the
presence of the object pictures and were instructed that the pictures were not part of the reward learning task.

To measure episodic memory formation for the incidental pictures, subjects completed a surprise memory probe. The memory probe took place immediately after reward learning (Study 1A) or approximately 24 hours later (Study 1B). For each picture, subjects indicated whether they thought that the picture was old (i.e. seen during the reward learning task) or new (i.e. not previously seen in this experimental context). Then subjects indicated how certain they were in their response on a 1-4 scale while viewing corresponding response options of “Guess”, “Somewhat certain”, “Very certain”, and “Completely certain”. Responses for the memory session were self-paced and multiple opportunities for rest breaks were included. Pictures were drawn from the 160 objects from the learning task and 80 new objects. Picture stimuli were presented without color overlay to avoid any bias in memory responses from choice color preferences from the reward learning task.

Additionally, in Study 1A, immediately following the reward learning task and before the memory probe, subjects completed a control reward learning task. Here, trial structure was the same, with the exception that no pictures of objects were presented in the colored choice options. Although behavior in the control task was based on more experience as it always followed the task with pictures, comparing behavior in the preceding task to the control task allows us to explore whether the presence of the incidental object pictures led to qualitative changes in reward learning behavior.

Following the memory probe session, subjects completed a written questionnaire which queried choice strategies and attention as well as whether subjects suspected a later
memory test of the incidentally presented pictures. Finally, subjects were informed how much money they had won in the experiment and were paid for their participation.

Study 2: Memory for feedback events. Procedures for Study 2 were similar to those for Study 1 with the critical difference that incidental trial-unique stimuli were presented during reward feedback instead of during choice (Figure 4.1B). As in Study 1, the experiment consisted of a reward task and a memory task, either performed on the same day (Study 2A) or performed on consecutive days (Study 2B).

During the reward learning task, choice options were represented by colored shapes (e.g., a yellow circle and triangle). On each trial, subjects had 2 s to choose between options (Figure 4.1B). After a 1 s inter-stimulus-interval, subjects received binary reward feedback for 1.75 s. A “win” was represented by a color overlaid on a unique picture stimulus; a “miss” was represented by a different color overlaid on a picture (Figure 4.1B). If no choice was recorded during the choice period, no reward outcome was displayed and "Too late!" was presented on the screen until the end of the trial. Trials were separated by variable duration ITIs (mean 3.25 s, range 1.0-13.5 s). The ITI fixation cross changed to black 1 s preceding the next trial to signal trial onset. The experiment consisted of 200 choices. As in Study 1, reward probabilities drifted gradually.

To investigate the influence of reward learning on episodic memory, trial-unique object pictures were included in the feedback. Subjects were informed of the presence of the object pictures but they were instructed that the pictures were not part of the reward learning task.
To allow the feedback picture memory probe to include color-overlaid objects without a reward-related response bias, color was counterbalanced. Specifically, subjects were instructed that the associations between color and feedback would switch after breaks in the game. The effect of this reversal on learning was minimized with task breaks that included text instructions followed by 10 “reminder” trials at the beginning of each mini-block of 50 trials. During these 10 reminder trials, text feedback of “Win!” or “Miss!” was presented above the associated color feedback square. Stimuli presented during these trials were repeated from the first 10 trials in the experiment and were not tested in the memory probe session. Thus, out of 200 trials total, 160 trials are of interest for subsequent memory analyses.

To measure episodic memory formation, subjects completed a surprise memory probe. The memory probe took place either immediately after reward learning (Study 2A) or approximately 24 hours later (Study 2B). The memory probe was as in Study 1 with the exception that pictures were presented in their original reward task color (Figure 4.1C). Subjects made memory judgments for each of the 160 pictures from in the reward learning task plus 80 additional novel pictures, in random order.

In Study 2, an error in data collection during the memory probe resulted in a loss of the last 40 of the 240 memory questions (including 26 old and 14 new pictures) for approximately half of subjects (12 of 21 subjects in Study 2A, 10 of 20 subjects in Study 2B). Data analysis was adjusted accordingly.

The task was presented using Matlab (Mathworks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997). The experiment was conducted on an Apple iMac computer.
Behavioral analysis. Analysis of the experimental data focused on the modulation of subsequent memory by ongoing reward learning. In all studies, we focused on the influence of reward on memory, and in Study 1, we also focused on the influence of choice on memory.

We first verified that participants adjusted their choices dynamically in response to reward feedback. Because of the fluctuating probability of reward, we could not estimate a learning curve or a percent correct over the course of the task. Instead, as in prior studies, a logistic regression model was fit to explain each participant’s sequence of choices in terms of two explanatory variables coding events from the previous trial: the choice made and whether it was rewarded (both coded as binary indicators) (Lau and Glimcher, 2005; Gershman et al., 2009; Daw et al., 2011; Li and Daw, 2011).

Next, we fit a temporal difference reinforcement learning (RL) model to subject’s choices (Sutton and Barto, 1998). This model was used to generate estimates of learning rate as well trial-by-trial variables to be used as predictors of memory formation. Such analyses attempt to explain the time series of choices in terms of previous events. (See O'Doherty et al., 2007; Daw, 2011 for reviews of the methodology.) The model tracks option value based on recent experience. Option values were initialized with a value of 0.5 and updated after reward feedback by the difference between the expected option value and the reward outcome – reward prediction error – scaled by the learning rate. Choices are presumed to be guided by a function of option value, with probabilities derived from a softmax distribution. The model also included a perseveration parameter, which captures a tendency to repeat or avoid selecting the same option on the subsequent
In Study 2, to eliminate possible contamination of model fitting by relearning of feedback-color associations when colors switched, we excluded the first 10 training choices and reset option value to 0.5 in each of the 4 mini-blocks.

Reinforcement learning model-derived memory predictors including value and reward prediction error were generated from a fixed-effects parameter fit across all subjects in Study 1 or Study 2. The fixed-effects fit was used because the low number of choices in the current experiment limits the reliability of individual parameter estimates (Daw, 2011). The RL model was also used to generate an “exploration” predictor. Choices can be categorized as exploratory (vs. exploitative) when a choice is not predicted by the values estimated by RL model, i.e., the chosen option has a lower model-predicted choice probability than the non-chosen option.

Memory prediction was estimated using random-effects logistic regression (implemented in STATA 9.1; StataCorp, College Station, TX). Memory, excluding missed reward learning trials, was coded as a binary outcome variable. All responses were included in reported analyses; removing low-confidence memory responses did not qualitatively alter the results. Reward was a primary predictor of interest across studies. In Study 1, choice was also a primary predictor of interest. A simple regression analysis was conducted for each study using reward and choice (Study 1) or reward (Study 2) as predictors. Additional regressions were conducted to test reward prediction error and the absolute value of reward prediction error (“surprise”) in place of reward.

A further regression analysis was conducted that included other predictors of interest as well as control variables. Additional predictors of interest included prior trial
reward, next trial reward, exploration, recent trial memory (mean over 4 preceding trials), future trial memory (mean over 4 subsequent trials), preceding inter-trial-interval, and subsequent inter-trial-interval. Control predictors included reward learning task trial number, memory probe trial number, reward learning reaction time, choice difficulty (negative entropy), and stimulus presentation side (for Study 1).

In Study 1, we also constructed a measure of the effect of individual picture stimuli on choice to use as a control variable in regressions. For each picture, across subjects we computed the probability that the subjects switched to or away from a given stimulus. During reward learning, these switches will be predominantly driven by prior feedback, but some may be due to particular stimuli being more likely to attract choice selection (or push choice selection away); we reasoned that averaging across subjects would capture some of the effect of individual stimuli. Each subject’s contribution to the mean was removed from the predictor.
Results

Reward learning

We verified that choice behavior was driven by recent rewards. In the reward learning task with incidental object pictures during choice (Study 1), reward on the prior choice predicted the next choice, and participants also exhibited the well-documented tendency to repeat (or perseverate), such that prior choice predicted the next choice (Table 5.1). In the reward learning task with pictures at feedback (Study 2), reward on the prior choice significantly influenced the subsequent choice, and prior choice also predicted the subsequent choice (Table 5.1). These effects of reward on choice are comparable to reward effects in similar experiments (e.g. Wimmer et al., 2012).

In Study 1A, we compared the effect of reward on choice between the task with incidental pictures and the control task with no pictures. In the control task, the effect of reward on the prior choice was significant (0.90 ± 0.06, P < 0.001) and not different than the effect in the task with pictures; the effect of prior choice was also similar (0.62 ± 0.06, P < 0.001). Thus, reward has a qualitatively similar influence on choice in both the task with incidental pictures and the task without pictures. This suggests that choice behavior was not strongly influenced by the visual saliency or preferences for the objects in the incidental pictures.

We also fit a simple reinforcement learning model to participant’s choices. For the reward learning task with incidental pictures during choice, across-participant fit parameters used in memory prediction were: learning rate = 0.61, beta = 1.97, and perseveration = 0.23. Mean parameter fits are reported in Table 5.1. For the control task without pictures in Study 1A, mean parameters were similar to the mean with pictures:
learning rate = 0.67 ± 0.07, beta = 5.75 ± 2.41, and perseveration = -0.01 ± 0.05.

Although changes in parameters may be driven by increased task experience, the decrease in learning rate from the reward learning task to the control task was significant ($t_{(19)} = 2.26, P < 0.05$). For Study 2, mean parameters are reported in Table 5.1. Fit parameters used in memory prediction were: learning rate = 0.53, beta = 2.07, and perseveration = 0.23.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Study 1: Choice Memory</th>
<th>Study 2: Feedback Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Delay</td>
</tr>
<tr>
<td>Reward</td>
<td>0.84 (0.06)</td>
<td>0.54 (0.05)</td>
</tr>
<tr>
<td>Choice</td>
<td>0.69 (0.06)</td>
<td>0.67 (0.05)</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.54 (0.08)</td>
<td>0.48 (0.06)</td>
</tr>
<tr>
<td>Beta</td>
<td>9.48 (3.20)</td>
<td>10.52 (3.06)</td>
</tr>
<tr>
<td>Perseveration</td>
<td>0.06 (0.04)</td>
<td>0.11 (0.05)</td>
</tr>
</tbody>
</table>

Table 4.1. Reward learning behavior.

Parameters for a logistic regression predicting choices, top (all P's < 0.001); parameters for a reinforcement learning model fit to choices, bottom.
Figure 4.2. Reward influence on memory.

Reward influence on memory for choice stimuli in Study 1 (left) and feedback stimuli in Study 2 (right). Reward enhances memory for both choice and feedback stimuli when tested after reward learning; when tested a day later, reward impairs memory for choice stimuli but strongly enhances memory for feedback stimuli. Reward memory effect coefficient values are derived from logistic regression. († P < 0.10, * P < 0.05, ** P < 0.01, *** P < 0.001)

Memory for choice events: Study 1

Study 1 was designed to probe the effect of ongoing reward learning on memory for stimuli experienced during choice. First, we asked whether participants remembered the pictures from the reward learning task. While the object pictures were incidental to the task, during a surprise memory test we found that participants nevertheless exhibited significant subsequent memory for the objects. When tested immediately (Study 1A), the mean corrected memory rate was $23.8\% \pm 4.1$ (hit rate, $56.7\% \pm 3.7$, false alarm rate, $32.9\% \pm 3.4$). When tested after a day (Study 1B), the mean corrected memory rate was $26.6\% \pm 3.3$ (hit rate, $57.5\% \pm 2.6$, false alarm rate, $30.9\% \pm 2.6$).
Next, we investigated whether memory was modulated by trial-by-trial experience during reward learning. When participants were given a surprise memory probe immediately, we found that reward increased memory ($t_{(18)} = 2.12, P < 0.05$; Study 1A, Figure 4.2). Surprisingly, however, when memory was tested after a delay in a separate group of participants, reward impaired memory ($t_{(27)} = -2.24, P < 0.05$; Study 1B, Figure 4.2) (immediate vs. delay reward effect: $t_{(47)} = 2.55, P < 0.05$).

We also investigated the influence of reward prediction error, the difference between expected and received reward, on memory. This was explored in a separate regression, as reward and reward prediction error are correlated (these variables are completely correlated when learning rate is zero). We found no effect of reward prediction error on memory at the immediate test, but we found that reward decreased memory after a delay ($t_{(27)} = -1.75, P < 0.10$). We found no effect of the absolute value of reward prediction error (surprise).

In addition to reward, we asked whether choice selection influenced memory for the stimuli presented during choice. Each reward learning choice included incidental stimuli in the chosen and non-chosen options. We found that memory was indeed strongly influenced by whether or not a picture was presented in a chosen option when tested immediately or after a delay (Figure 4.3). The memory benefit of chosen vs. non-chosen items was approximately 9% (Study 1A, 10% ± 1.8, Study 1B, 8.5% ± 1.5).
Figure 4.3. Choice influence on memory.

Chosen stimuli are remembered significantly more often than non-chosen stimuli in Study 1, regardless of delay. (*** \( P < 0.001 \).)

The modulation of memory by reward was not different for chosen and non-chosen stimuli. In separate analyses for chosen and non-chosen options, at the immediate memory test, we found positive effects of reward on memory for both stimuli (chosen, \( t_{(18)} = 1.05, \text{n.s.} \); non-chosen, \( t_{(18)} = 1.68, P < 0.10 \)). In the delayed memory test, we found negative but non-significant effects of reward for both stimuli (chosen, \( t_{(27)} = -1.61, P = 0.11 \); non-chosen, \( t_{(27)} = -1.67, P < 0.10 \)).

While the effect of reward on memory was similar for chosen and non-chosen stimuli, the effect of reward on memory may differ because of changes in choice behavior. We thus investigated the effect of choice switches on memory and the reward modulation of memory. Participants switched their choice selection from the preceding trial (e.g. from green to blue) on approximately a third of all trials (Study 1A, 30.8% ± 1.8; Study 1B, 32.1 ± 2.7). Switches had no main effect on memory. For the effect of
reward on memory, when memory was tested immediately, reward increased memory for items on switch and non-switch choices (P’s <0.15). However, choice switching had a large influence on the negative reward effect on memory at a delay: reward was a significant negative predictor of memory only for stimuli presented on switch choices (switch: -0.191 ± 0.057, t(27) = -3.37, P < 0.001; non-switch: -0.211 ± 0.038, t(27) = -0.55, n.s.; this effect was found for chosen and non-chosen stimuli).

These data demonstrate that reward learning modulates episodic memory formation for stimuli incidentally experienced during choice. Reward initially increased memory, but when tested after a delay of approximately 24 hours, reward decreased memory. Choice itself strongly increased memory formation, regardless of delay.

**Memory for feedback events: Study 2**

Study 2 was designed to probe the effect of ongoing reward learning on memory for stimuli experienced during feedback; this is in contrast to Study 1, where incidental pictures were presented during choice. First, we asked whether participants remembered the pictures from the reward learning task. While the object pictures were incidental to the task, a surprise memory test revealed that participants exhibited significant subsequent memory for the objects. When tested immediately (Study 2A), corrected hit rate was 30.4% ± 2.9 (hit rate: 56.0 % ± 3.0, false alarm rate: 25.6% ± 2.3). When tested after a delay of approximately 24 hours (Study 2B), corrected hit rate was 24.9% ± 2.8 (hit rate: 48.3% ± 3.3, false alarm rate: 23.3% ± 2.8; corrected hit rate was not significantly different between Study 2A and 2B).
Next, we investigated whether memory was modulated by trial-by-trial experience during reward learning. We found that when tested immediately, there was a trend for a positive effect of reward on memory ($t_{(19)} = 1.66$, $P = 0.097$; Figure 4.2). When tested after a delay, reward was a significant positive predictor of memory ($t_{(19)} = 3.55$, $P < 0.001$; Figure 4.2). Although numerically larger, the effect of reward on memory was not significantly greater in the delayed memory test condition. In a separate analysis, we found that reward prediction error was a positive but weaker predictor of memory than reward (immediate test, $t_{(18)} = 1.58$, $P < 0.15$; delayed test $t_{(18)} = 2.28$, $P < 0.05$). We found no effect of the absolute value of reward prediction error (surprise).

<table>
<thead>
<tr>
<th>Immediate probe</th>
<th>Delayed probe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
</tr>
<tr>
<td>Choice</td>
<td>5.93***</td>
</tr>
<tr>
<td>Reward</td>
<td>2.12*</td>
</tr>
<tr>
<td>constant</td>
<td>8.60***</td>
</tr>
</tbody>
</table>

Table 4.2. Study 1: Choice stimulus memory prediction.

Choice and reward modulation of memory for stimuli seen during choice when tested immediately and after a delay. (* $P < 0.05$, ** $P < 0.01$ *** $P < 0.001$.)
Table 4.3. Study 2: Feedback stimulus memory prediction.

<table>
<thead>
<tr>
<th></th>
<th>Immediate probe</th>
<th></th>
<th>Delayed probe</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>coef. (se)</td>
<td>t</td>
<td>coef. (se)</td>
</tr>
<tr>
<td>Reward</td>
<td>1.66†</td>
<td>0.129 (0.078)</td>
<td>3.55***</td>
<td>0.276 (0.078)</td>
</tr>
<tr>
<td>constant</td>
<td>1.21</td>
<td>0.104 (0.086)</td>
<td>-2.44</td>
<td>-0.285 (0.117)</td>
</tr>
</tbody>
</table>

Reward modulation of memory for stimuli seen during feedback when tested immediately and after a delay. († P < 0.10, * P < 0.05, ** P < 0.01 *** P < 0.001.)

Individual Differences

Finally, we explored whether individual differences in reward learning behavior predicted memory formation. For pictures presented during choice (Study 1), we found that memory response hit rate was significantly correlated with how well participant’s choices were described by a reinforcement learning model (r = 0.43, P < 0.01; n = 49), such that participants with a better model fit had higher recognition hit rates. For pictures presented during feedback (Study 2), we found the opposite relationship: memory response hit rate was negatively correlated with reinforcement learning model fit (r = -0.42, P < 0.01; n = 40; difference in correlations, P < 0.001). These opposing correlations suggest that the relationship between episodic memory encoding and reward-guided choice depends on whether memory processes are engaged during choice or feedback.
Discussion

Encoding of episodic experience and learning from feedback are often intermixed in everyday life. However, we know very little about how reward and episodic memory interact. To explore whether and how episodic memory formation is affected by ongoing reward learning, we adapted a basic reinforcement learning game to include incidental trial-unique object stimuli during choice selection or during reward feedback. We found that when memory was probed immediately after reward learning, reward increased memory for choice stimuli and marginally increased memory for feedback stimuli. Surprisingly, after a delay of one day, we found that memory for choice and feedback events was oppositely modulated by reward: reward decreased memory for choice stimuli but enhanced memory for stimuli presented during reward feedback. Choice itself was a positive modulator of memory when tested immediately and after a delay.

Reward and memory.

The beneficial effect of reward on episodic memory for feedback events fits well with prior research on reward and memory where learning was confined to single trials (Wittmann et al., 2005; Adcock et al., 2006). In our experiment, reward is experienced within a dynamic learning game, and thus the modulation of memory by reward that we observe occurs at the same time that reward is updating choice values in the task. In prior research, potential reward value was indicated before stimulus presentation (Adcock et al., 2006) or was cued by the category of stimulus presentation (Wittmann et al., 2005); our results additionally demonstrate that reward can increase memory for stimuli that are presented during reward feedback itself. Additionally, in our experiment we are able to
compare the effects of surprising positive vs. negative feedback on memory. Surprising feedback may increase arousal, which has been found in studies of emotion and memory to increase recollection (McGaugh, 2004; Phelps, 2004). However, we did not find an influence of overall feedback surprise on memory for choice or feedback events.

The reward enhancement of memory for feedback episodes is in line with cellular research on the enhancing effects of dopamine release on memory (Shohamy and Adcock, 2010; Lisman et al., 2011). In the reward learning task, reward was never fully predicted, and thus unexpected reward feedback likely led to dopamine release in the striatum (Roitman et al., 2008) and possibly the hippocampus. Interestingly, the positive effect of reward on feedback memory was numerically higher after an approximately 24-hour delay between learning and test, suggesting that neural processes such as consolidation enhance the effect of reward on memory. Such a consolidation benefit of reward-related memory is supported by research on plasticity in the hippocampus, where dopamine release during learning leads to stronger memory after consolidation (Huang and Kandel, 1995; Bethus et al., 2010).

Unexpectedly, episodic memory for choice stimuli was negatively affected by reward after a delay, in contrast to the positive enhancement observed for feedback memory. The origin of this dissociation between choice memory and feedback memory is unknown; if the same reward signal from feedback events updates the value of choices, it is unclear why memory for choice events would be modulated negatively by reward.

The different effects of reward on choice and feedback stimuli may be related to cognitive processes associated with ongoing behavioral adjustment in the reward learning task. Reward or miss feedback is relatively simple to process and does not require any
immediate action. The effect of reward feedback on choice options is more complex, because feedback must be used to update the value of options that are no longer present on the screen. This may influence the effect of reward on memory for choices. Another possibility is that the temporal separation between choice and feedback explains the dissociable effects of reward on memory: reward could decrease memory for stimuli further from feedback while reward could increase memory for stimuli closer to feedback.

Changes in choices over the task provide further insight into how choice memories are negatively modulated by reward after a delay. The negative effect of reward on choice memory was only observed for choice memory when participants switched choices from the preceding trial. Switching commonly happens when a previously selected option does not lead to reward as often as expected or when participants explore the other option. The influence of switching on memory was only evident after a delay; when tested immediately, we found a positive effect of reward on memory regardless of switching. The different effects of reward on memory for choice and feedback events is an interesting area to explore in future research in choice situations as well as in reward learning tasks that do not involve choice (for example, in Pavlovian learning).

The dissociation between the effect of reward on choice and feedback memory may have important implications for decision making. When decision-related stimuli are encountered later, our results suggest that prior positive experience will decrease recognition of choice-related stimuli. In contrast, prior positive experience will increase the recognition of feedback-related stimuli. In our study, this dissociation in the effect of
reward on memory was found when events were separated by several seconds. The effect on memory and future behavior may be even more significant if this dissociation extends to situations with longer gaps between decision and feedback.

In our experiment, we may have been able to detect the dissociation between reward effects on choice and feedback memory because choice options were not present during feedback. This suggests that studies that do not separate choice and feedback may obtain weak or null effects of reward. However, in many experiences, stimuli present during choice are also present during feedback; for example, when exploring a new restaurant, the environment will be present from the time of ordering to evaluation. Our results make no clear predictions about reward and memory in these kinds of situations where events persist from decision to feedback. After a delay, memory could be dominated by the negative reward effect for choice stimuli, the positive reward effect of feedback stimuli, or a combination of both.

Choice and memory.

In addition to reward, choices during reward learning strongly increased memory for incidental stimuli presented within choices. Previous studies have not investigated episodic memory during a decision making task, and thus it was unknown whether choice would influence memory. The positive effect of choice on memory may be related to the greater value participants assign to chosen vs. non-chosen items. Participants’ goal was to choose the option that they currently thought was most likely to lead to reward, and thus choice itself may be considered a proxy for participant’s internal value estimates.
In an exploratory analysis, value, as estimated by a reinforcement learning model, exhibited some evidence for modulation of choice picture memory. When tested immediately, both chosen value and non-chosen value increased episodic memory at a trend level (P’s < 0.10). When tested after a delay, only non-chosen value had a positive effect on memory (P < 0.05). In future research, it will be of interest to explore connections between a memory benefit for choices and choice value signals, such as those commonly found in the ventromedial prefrontal cortex (Daw et al., 2006; Hare et al., 2008).

Improved memory for choices is also likely related to attention. Recent research investigating visual fixation during product decision making has found that participants are biased to look at higher valued chosen options (Krajbich et al., 2010). Increased visual attention would have clear benefits for memory encoding. In the current experiment, we did not record measures of eye movements, but we would predict that participants fixated more on the chosen than non-chosen option. Additionally, choice selection may elicit similar memory benefits that have been reported for encoding during self-directed exploration (Voss et al., 2011).

Conclusion and future directions.

Our results demonstrate that experiences during ongoing reward learning influence episodic memory. The modulation of memory by experience may be particularly important for future behavior. For example, episodic memories formed during reward learning may influence future decisions. In future research, it will be of interest to explore how choice and reward feedback may bias not only later recollection
but also value-based decision making. While the reward learning task in our experiments looked at learning over a short time horizon, it may be possible for episodic memory and reward associations to bias choices hours and days later.

One interesting question for future research is whether memory also contains traces of the reward feedback experienced during the choice trial. It is unknown if fleeting reward associations from only a single experience are encoded in memory. Some evidence suggests that reward associations are indeed remembered, at least when stimuli are repeated multiple times (Eppinger et al., 2010). Neurally, a recent functional imaging study of memory suggests that traces related to initial reward associations may be detectable upon re-exposure (Kuhl et al., 2010). Future research could determine whether participants can explicitly recall the reward feedback associated with choice or feedback events.

In conclusion, we find that during reward learning, reward feedback does not just bias ongoing choices: reward exerts an additional influence on episodic memory formation. Additionally, choice strongly increases memory. Together, these findings support the idea that the brain encodes more than just the value of choices and feedback during learning, suggesting that episodic memory and reward are not isolated processes.
Chapter 5: Choice and reward modulate episodic memory in the human brain

Chapter 5 is from an in preparation manuscript: Wimmer, G.E., Braun, E.K., and Shohamy, D.
In Chapter 4, I demonstrated that choice selection enhances memory. Additionally, reward enhances memory for feedback events, but has a shifting effect for choice events such that reward initially enhances memory but decreases memory after a delay. Here, I explore neural correlates of the enhancing effects of choice on memory and the interaction of reward feedback and memory, focusing on the long-term effects on memory using a 24-hour delay between reward learning and memory.

Abstract

Memory for events is critical for adaptive behavior, but because memory is limited, it is important to selectively encode important events. Recent research has begun to investigate how episodic memory is modulated by salient experiences such as reward, but surprisingly little is known about how memory is affected by ongoing learning and decision making. To explore how memory is influenced by learning, we employed functional MRI and a reward learning task. Participants made choices between two options (colored squares), each associated with a drifting reward probability. Incidental, trial-unique object pictures were overlaid on each option. One day after the reward learning task, participants were given a surprise memory probe for the pictures. Although unrelated to the task, we found that participants had significant memory for the pictures. Neurally, activation in the ventromedial prefrontal cortex was related to choice value as well as to successful memory formation for pictures in chosen options. A similar effect of choice memory was found in the hippocampus: activation in the hippocampus was related not just to overall memory, but specifically to the effect of choices on memory. In separate regions of the striatum, reward prediction error signals at feedback interacted both negatively and positively with memory. Further, activity in the striatum correlated
with individual differences in the behavioral influence of reward on memory. These results demonstrate that episodic memory formation is influenced by ongoing learning and provide insight into how separate neural systems for episodic memory and reward learning are modulated by both processes.
Introduction

We cannot remember everything – our memory formation is selective. Prior decisions and outcomes may bias what we remember and what we forget; this selective memory, in turn, determines the knowledge we can use in future decisions. Memory and learning from feedback are often closely related during dynamic behavior such as goal pursuit, but experimentally, memory formation is often studies in isolated environments. The hippocampus is a critical part of a neural system for encoding specific experiences, termed relational memory (Davachi, 2006). Recently, studies have begun to investigate how memory and hippocampal activity are modulated by salient aspects of the environment, such as reward (Wittmann et al., 2005; Adcock et al., 2006; Mather and Schoeke, 2011). Nevertheless, it remains largely unknown whether and how episodic memory is modulated by ongoing behavior, such as learning from reward feedback. Is memory affected by the value participants assign to choices? Is memory strengthened by the same neural mechanisms by which reward feedback updates values?

While memory for everyday experiences and events often takes place alongside active exploration and goal seeking, few studies have probed the effect of exploration and choice on memory. Recently, a benefit for volitional exploration on memory formation in an intentional encoding paradigm was demonstrated, an effect supported by increased hippocampal-cortical coupling during encoding (Voss et al., 2011).

If active choice does modulate episodic memory, this effect may be related to neural correlates of predicted reward value. In human fMRI studies, choice value has been shown to correlate with activation in the ventromedial prefrontal cortex (VMPFC) (Daw et al., 2006; Knutson et al., 2007; Plassmann et al., 2007; Hare et al., 2008).
Participant’s choices can also be predicted by attention, as measured by visual fixations (Krajbich et al., 2010). This attentional effect has been related to valuation in the VMPFC (Lim et al., 2011).

Reward may also modulate episodic memory. The reward value of cues has been shown to enhance incidental memory (Wittmann et al., 2005; Mather and Schoeke, 2011), an effect related to midbrain and hippocampal activity (Wittmann et al., 2005). In a reward-motivated encoding context, reward has also been shown to enhance memory and hippocampal activation (Adcock et al., 2006). In addition to reward, many previous studies have explored the influence of emotion on memory. Emotion, in particular arousal, has also been shown to increase memory (McGaugh, 2004; Phelps, 2004).

Collectively, these findings provide preliminary evidence for the possible role of choice and reward on episodic memory formation. However, other research suggests that memory may be unaffected by reward learning. This hypothesis is suggested by results showing a competition between an episodic memory system in the hippocampus and a reward learning system in the striatum (Poldrack and Packard, 2003). Similarly, reward learning may have no effect on memory formation simply because reward feedback only affect the value of options, to the exclusion of modulating episodic memory formation.

Our goal was to better understand how choice and reward modulate memory and to explore the neural systems that predict memory during learning. We adapted a well-established reward learning task that has been shown in previous studies to reliably elicit neural activation related to choice value and reward prediction error, the difference between choice value and reward feedback (Daw et al., 2006; Schönberg et al., 2007). Participants attempted to choose the most rewarding option in a simple learning game.
with fluctuating reward probabilities. Choice options also contained incidental object pictures. Functional magnetic resonance imaging (fMRI) data were collected during reward learning. Memory was measured in a surprise memory probe approximately 24 hours after encoding. Our paradigm brings together learning and decision making in the kind of dynamic environment experienced in everyday behavior, where there are no discrete episodic learning phases and reward learning phases, and where multiple memory systems, including reward learning in the striatum and episodic memory formation in the hippocampus, may be engaged at the same time.
Figure 5.1. Reward learning and memory task.

A, Reward learning task where participants tried to maximize reward by making repeated choices between colored options. B, Choice options included incidental trial-unique picture stimuli that were unrelated to the reward learning game. Surprise subsequent memory test, administered a day after the reward learning session.

Methods

Participants. Thirty-two fluent English speakers with no neurological or psychiatric disorders and normal or corrected-to-normal vision participated in the study. Informed consent was obtained in a manner approved by the Columbia University Institutional Review Board. Two subject’s data were excluded due to errors in behavioral data collection, leaving thirty subjects (15 female; mean age, 23 years; range, 18-35). For the
reward learning and subsequent memory MRI analyses, we first analyzed all subjects. In later analyses, we focused on a subgroup of 18 subjects with higher memory performance (> 5% corrected memory). Subjects were paid $20 per hour for the approximately 3.5-hour duration of participation in the two sessions plus one-half of the nominal rewards they earned in the experimental task.

Task. The experiment consisted of two sessions performed on consecutive days. On day one, subjects completed a reward learning task that included incidental exposure to trial-unique picture stimuli. On day two, subjects returned to the lab to complete a surprise probe of memory for pictures that were seen on day one.

In the reward learning task (Figure 5.1A), subjects chose one of two presented options and then received monetary feedback. The reward learning task is a variant of a “two-armed bandit” task (e.g., Schönberg et al., 2007), with the two choice options distinguished by a (blue or green) colored frame around a central square. The left-right location of the options was permuted randomly from trial to trial. On each trial, subjects had 2 s to choose between the options, using an MR-compatible button response pad held in the right hand (Figure 5.1A). So as to maintain equal visual saliency of both options and the included pictures, the choice selection was not highlighted. Options remained on the screen for 2.5 s. After a blank 1 s inter-stimulus-interval, subjects then received binary reward feedback for 1.75 s. A $0.25 “win” outcome was represented by an image of a quarter dollar (Figure 5.1A) and a $0.00 “miss” outcome was represented by a phase-scrambled image of a quarter dollar. If no choice was recorded during the choice period, no reward outcome was displayed and "Too late!" was presented on the screen.
until the end of the trial. A jittered fixation inter-trial-interval (ITI) preceded the next trial (mean 3.5 s, range 1.5-13.5); to signal trial onset, the fixation cross shifted from white to black 1 s preceding the next trial.

Subjects were instructed that each option was associated with a different probability of reward, that these probabilities could change slowly, and that their goal was to attempt to find the most rewarding option at a given time in order to earn the most money. They were also instructed that rewards were tied to the color and not the option position. Subjects completed a short paid practice version to familiarize them with the task, and this practice was repeated once participants were situated inside the MRI scanner.

Each of the two options was associated with a different probability of monetary reward. Throughout the 100 choices in the experiment, the reward probabilities diffused gradually to encourage continual learning. Reward probabilities were generated using Gaussian random walks with reflecting boundary conditions at 20% and 80%. Two instantiations of random walks were generated of mean 50%; these were inverted to create four total reward probability sequences. Reward probability sequences were counterbalanced between subjects, as was the mapping of color assignment to the underlying reward sequences.

To investigate the influence of reward learning on memory formation, trial-unique object pictures were included in the center of each option. Subjects were informed of the presence of the object pictures but they were instructed that the pictures were not part of the reward learning task. Four different lists of 200 stimuli were selected for the reward
learning task, and 100 stimuli were reserved as new items for the memory probe. Object picture lists were counterbalanced between subjects.

Following the initial reward learning task, subjects completed a control reward learning task with 100 trials that lacked object pictures. In this task, choice options were represented by shapes (e.g., a yellow circle and triangle) instead of colored frames.

The duration and distribution of null events was optimized for estimation of rapid event-related fMRI responses as calculated using Optseq software (http://surfer.nmr.mgh.harvard.edu/optseq/). The task was presented using Matlab (Mathworks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997). The reward learning task was projected onto a mirror above the subject’s eyes.

On the day after the first session, subjects returned to the lab and completed a surprise memory probe to measure subsequent memory for the object pictures incidentally presented during the reward learning task. In the memory session, subjects saw each of the 200 pictures that had been presented on the preceding day plus 100 novel pictures, in a pseudo-randomized order. Pictures were presented in the center of the screen above the response options “Old” and “New” (Figure 5.1B). Pictures were surrounded by a grey frame to avoid the influence of color choice biases from reward learning on memory responses. Subjects indicated by key press whether they thought that the picture was old (seen during the preceding day’s reward learning task) or new (not previously seen in this experimental context). Then, subjects indicated how certain they were in their response on a 1-4 scale while viewing corresponding response options of “Guess”, “Somewhat certain”, “Very certain”, and “Completely certain”. Responses for
the memory session were self-paced. Multiple opportunities for rest breaks were included. The memory probe session was completed on a laptop computer.

Following the memory probe session, subjects completed a written questionnaire that queried choice strategies and attention during the reward task and memory probe session, as well as whether they thought during the reward task that their memory may be tested for the incidentally-presented object pictures. Finally, subjects were informed how much money they had won in the experiment and were paid for their participation.

**Imaging procedure.** Whole-brain imaging was conducted on a 3.0T Phillips MRI system at Columbia University’s Program for Imaging and Cognitive Sciences, using a SENSE head coil. Head padding was used to minimize head motion; no subject’s motion exceeded 2 mm in any direction from one volume acquisition to the next. Functional images were collected using a gradient echo T2*-weighted echoplanar (EPI) sequence with blood oxygenation level-dependent (BOLD) contrast (TR = 2000 ms, TE = 15 ms, flip angle = 82, 3 X 3 X 3 mm voxel size; 45 contiguous axial slices). For each functional scanning run, five discarded volumes were collected prior to the first trial to allow for magnetic field equilibration. Four functional runs of 232 TRs (7 min and 44 s) were collected, each including 50 trials. Following the functional runs, structural images were collected using a high-resolution T1-weighted MPRAGE pulse sequence (1 X 1 X 1 mm voxel size).

**Behavioral analysis.**
First, we verified that subjects adjusted their choices dynamically in response to the rewarding outcomes in both the reward learning task with incidental stimuli and in the second control task without pictures. Because of the fluctuating probability of reward, we could not estimate a learning curve or a percent correct over the course of the task. Instead, as in prior studies, a logistic regression model was fit to explain each subject’s sequence of choices in terms of two explanatory variables coding events from the previous trial: the choice made and whether it was rewarded (both coded as binary indicators) (Lau and Glimcher, 2005; Gershman et al., 2009; Daw et al., 2011; Li and Daw, 2011).

To estimate reward learning rates in the primary task and the control task, as well as to generate trial-by-trial reward prediction error predictors of memory, we fit a temporal difference reinforcement learning (RL) model to subject’s choice behavior (Sutton and Barto, 1998). Such analyses attempt to explain the timeseries of choices in terms of previous events. (See O’Doherty et al., 2007; Daw, 2011 for reviews of the methodology.) Option values were initialized with a value of 0.5 and updated after reward feedback by the difference between the expected option value and the reward outcome – reward prediction error – scaled by the learning rate. Choices are presumed to be guided by a function of option value, with probabilities derived from a softmax distribution. The model also included a perseveration parameter, which captures a tendency to repeat or avoid selecting the same option on the subsequent trial (Lau and Glimcher, 2005; Schönberg et al., 2007).

Parameters were optimized for each subject using an optimization routine which included 20 starting points to avoid local minima. Median values for the learning rate,
softmax inverse temperature (beta), and perseveration were used to generate choice values and reward prediction errors. Median values were used as the relatively low number of choices in the current experiment (selected to maximize subsequent memory performance) limits the reliability of individual parameter estimates.

To predict memory formation during reward learning, we employed random-effects logistic regression (implemented in STATA 9.1; StataCorp, College Station, TX). The model predicted memory for incidental pictures (a binary outcome variable) via trial-by-trial variables from the reward learning task. All responses were included in reported analyses; removing low-confidence memory responses did not qualitatively alter the results. In the memory analyses, we constructed two models. A simple model included only choice and the effect of feedback, formalized as reward prediction error.

A complete memory prediction model also included decision reaction time, reward learning trial number, memory probe trial number, prior trial reward, next trial reward, choice difficulty (negative entropy), exploratory choices, recent memory (mean over the past four trials), future memory (mean over the next four trials), the likelihood that a stimulus would lead to a switch in choice (derived from the group’s behavior), and the side of stimulus presentation. Results for predictors of interest do not change with the inclusion of these additional control variables. Additionally, we hypothesized that reward feedback may have a larger influence on memory when reward and omitted reward is more surprising. Thus, we conducted a second regression model that only included trials where the absolute value of reward prediction error was greater than or equal to the mean (0.5).
Imaging analysis. Preprocessing and data analysis was performed using AFNI (Cox, 1996) and Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were coregistered manually using AFNI; the remainder of the analysis was completed in SPM. Images were realigned to correct for subject motion and then spatially normalized by estimating a warping to template space from each subject’s anatomical image and applying the resulting transformation to the EPIs. Images were resampled to 2 mm cubic voxels, smoothed with an 8 mm FWHM Gaussian kernel, and filtered with a 128 s high-pass filter. Data collected for one subject during the 2011 Virginia earthquake were inspected with independent component analysis as implemented in FSL’s MELODIC (Beckmann and Smith, 2004) to ensure that no artifacts were introduced by the earthquake.

For analysis of all phases of the experiment, fMRI model regressors were convolved with the canonical hemodynamic response function and entered into a general linear model (GLM) of each subject’s fMRI data. The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the GLM to account for residual effects of subject movement. Linear contrasts of the resulting SPMs were taken to a group-level (random-effects) analysis. We report results corrected for family-wise error (FWE) due to multiple comparisons (Friston et al., 1993); this approach assesses the strength of activations defined by an initial and arbitrary uncorrected threshold, which we take as $P < 0.005$ for all analyses. Accordingly, for display purposes, we render all SVC significant activations at this threshold. We conduct this correction at the peak-level within small volumes for which we had an a priori hypothesis or at the
whole-brain cluster level. For regions of interest in the striatum and in the MTL (including both the hippocampus and parahippocampal cortex) we used anatomically defined masks derived from the AAL atlas (Tzourio-Mazoyer et al., 2002). For more focused a priori analyses in the hippocampus, we used 6mm spherical regions of interest centered at (-20, -10, -18) and (20, -12, -18) based on a previous report of reward and memory interactions (Adcock et al., 2006). Further, choice value and reward prediction error activation during the reward learning task was used to define masks for small volume correction in the VMPFC and striatum, respectively (thresholded at P < 0.005, uncorrected). All voxel locations are reported in Montreal Neurological Institute (MNI) coordinates, and results are displayed overlaid on the average of all subjects’ normalized high-resolution structural images.

In an initial general linear model (GLM), we estimated BOLD correlates of trial-to-trial reinforcement learning variables. The model included regressors at choice for choice probability (referred to henceforth as choice value) and at feedback for reward prediction error (both modeled for 0.5 s). This model was estimated for functional data from the reward learning task with pictures, the control task without pictures, and both tasks combined. Mean RL parameter values across subjects were used to generate regressors for the fMRI model, as individual parameter fits in tasks and models of this sort tend to be noisy, and regularization of the fit parameters across subjects tends to improve a model’s subsequent fit to fMRI data (following previous work: Daw et al., 2006; Schönberg et al., 2007; Gershman et al., 2009).

Our primary GLM analysis targeted neural signals predictive of subsequent memory for the incidentally presented object pictures. This model included two
regressors during the choice period for memory for chosen items and non-chosen items (modeled for 0.5 s) and two regressors during the feedback period for reward prediction error modulated by memory for chosen items and non-chosen items (modeled for 0.5 s).

At the second level, we constructed contrasts of interest of the resulting beta maps: at choice, the contrasts of interest were overall memory, memory for chosen items, and memory for non-chosen items; at reward feedback, the contrast of interest was the modulation of reward prediction error by memory. Finally, individual differences in the influence of reward prediction error on memory were correlated with the beta coefficients from the interaction of reward prediction error and memory. To capture the full range of individual variability and to maximize power to detect an effect, this analysis was conducted in the full group of participants.
Results

Behavioral Results

First, we established that participants used reward feedback to guide their choices in the reward learning task. In a logistic regression model of choice, the effect of reward on the prior choice was significant (0.766 ± 0.043 (mean ± SEM), P < 0.001), as was the effect of prior choice (0.731 ± 0.043, P < 0.001). In the control task with no pictures, the effect of reward on the prior choice was also significant (0.994 ± 0.047, P < 0.001), as was the effect of prior choice (0.924 ± 0.047, P < 0.001). These effect did not differ between the two tasks. The similar influence of reward on choice from these two tasks indicates that reward learning behavior was not qualitatively changed by the inclusion of incidental pictures in choice options.

Using a reinforcement learning model, mean parameters were: learning rate = 0.62 ± 0.05, beta = 2.94 ± 0.39, and perseveration = 0.09 ± 0.05. In the control task, mean parameters were: learning rate = 0.71 ± 0.05, beta = 3.69 ± 0.52, and perseveration = 0.05 ± 0.04. While any differences between these parameters could be due to additional experience in the reward learning task, perseveration was lower in the control task (t(29) = 2.30, P < 0.05).

Next, we turned to our primary question of whether and how episodic memory is modulated by reward learning. Although the object pictures were incidental to the participant’s goal of maximizing their reward, participants nevertheless exhibited significant subsequent memory. Corrected hit rate was 16.9% ± 2.2 (hit rate: 53.0% ± 2.4; false alarm rate: 36.1% ± 2.8 (n = 18)). We found that memory was strongly
enhanced when a picture was presented in a chosen option (0.110 ± 0.035; \( t_{(16)} = 3.16, P < 0.01 \); choice memory benefit = 5.4%; Figure 5.2).

We then explored the modulation of memory by reward feedback, focusing on reward prediction error, a trial-by-trial variable computed as the difference between expected reward and received reward. The effect of reward prediction error on memory was negative but non-significant (-0.064 ± 0.064; \( t_{(16)} = -1.00 \); Figure 5.2). As more surprising reward feedback may have a stronger effect on memory, we next restricted our analysis to trials with high positive and negative reward prediction errors. We found a significant negative effect of reward prediction error on memory (-0.265 ± 0.090; \( t_{(16)} = -2.94, P < 0.01 \)). We found no effect of feedback “surprise”, the absolute value of reward prediction error, on memory. Thus, surprising positive reward feedback led to a decreased likelihood of memory formation, while surprising miss feedback led to an increased likelihood of memory formation. The finding of a negative influence of reward on memory after a day of consolidation replicates a result in a separate behavioral study (Chapter 4).
Influence of choice and reward feedback on the likelihood of successful memory formation. (* P < 0.01.)

**fMRI Results**

We focused our imaging analyses on detecting brain activity related to the modulation of episodic memory by reward learning. First, we found that overall memory was predicted by activation in the posterior hippocampus (Figure S5.1) and visual cortex during choices (Figure S5.2). Next, we localized regions correlated with choice value and reward prediction error. Replicating prior results, we found that activity in the VMPFC correlated with chosen option value ($z = 4.78 \ (-6, 50, -8)$; Figure S5.3) and that activity in the striatum correlated with reward prediction errors at feedback (right $z = 4.87 \ (30, -10, -6)$, left $z = 4.21 \ (-30, -10, 6)$; Figure S5.3).

*Neural correlates of the benefit of choice on memory.* In the region of the VMPFC correlated with choice value, we found a significant subsequent memory effect for chosen, but not non-chosen, items (Figure 5.3A). Greater activity correlated with
memory for chosen items. The same region of the VMPFC region was responsive to value and choice memory, but the value signal was not modulated by memory. Additionally, we found that activation in the bilateral anterior hippocampus correlated with memory for chosen stimuli (Figure 5.3B).

![Figure 5.3](image_url)

**Figure 5.3. Activation predicting memory for chosen options.**

Activation correlates with chosen option memory in the VMPFC (z = 3.56 (-6 46 -18)), A, and hippocampus (left z = 3.17 (-18, -10, -20), right z = 3.48 (18, -10, -18)), B. (Activations survive P < 0.05 SVC; all MRI figures P < 0.005 uncorrected for display.)
Interactions between memory formation and reward prediction error. We next explored activation to reward feedback that may be modulated by memory formation. Thus, we examined the interaction of subsequent memory and reward prediction error. Regions where activity significantly correlates with the interaction show differential prediction error activity for remembered versus forgotten trials. Neural effects were similar for chosen vs. non-chosen items, so results are presented for the combined interaction. We first examined this interaction in the striatal region with the strongest effect of reward prediction error, the ventral putamen, (Figure S5.3). Activation in the right ventral putamen negatively correlated with the interaction of episodic memory and reward prediction error (Figure 5.4). In this region, activity was significantly correlated with prediction error only for trials with forgotten stimuli.
Figure 5.4. Memory and prediction error interaction in the right putamen.

(z = 3.30 (28, -6, -4); P < 0.05 SVC in a putamen region of interest mask derived from reward prediction error responses.)
We next explored the interaction of memory and reward prediction error in the striatum more broadly. In the right caudate, activation was significantly positively correlated with the interaction of reward prediction error and memory (Figure 5.5). Thus, in contrast to the effect in the putamen, in the caudate, the correlation with reward prediction error was greater on remembered trials.

We also probed interactions between memory and reward prediction error in the medial temporal lobe. In the left hippocampus, we found a positive interaction of reward
prediction error and memory (-24, -14, -18, z = 2.80, P < 0.05 SVC). This cluster was adjacent to the region of the left hippocampus that correlated with memory for chosen items.

Finally, we explored whether individual differences in the negative effect of reward on memory related to neural activity during learning. Reward prediction error tended to negatively influence memory; we correlated individual differences in this behavioral effect with the neural interaction of memory and reward prediction error. We found a significant correlation between participant’s tendency to remember stimuli on rewarded trials and activity in the caudate correlated with interaction of reward prediction error and memory (Figure 5.6). While reward prediction error tended to negatively influence memory formation, in the caudate, only participants with a beneficial effect of reward on memory exhibited differential prediction error activity for remembered vs. forgotten items.
Figure 5.6. Reward memory effect correlates with caudate interaction of prediction error and memory.

\( z = 3.69 \) (14, 20, 8); \( P < 0.05 \) SVC in a region of interest based on the striatal interaction of memory and reward prediction error. Reward memory measures were derived from each participant’s memory regression coefficient for reward prediction error. Illustration of correlation in striatal region independently identified by the interaction \( r = 0.535, P < 0.01 \), bottom.
Discussion

We explored the modulation of episodic memory formation by ongoing reward learning. Our results demonstrate that episodic memory for incidental pictures is influenced by choice and reward. Neurally, memory for pictures in chosen options was related to activation in the VMPFC, a region also correlated with choice value. Additionally, prediction error responses in the striatum were significantly modulated by successful memory formation. While it may be predicted that memory would be unaffected by reward learning, our results support a perspective suggesting a cooperative interaction between memory systems in the hippocampus and striatum (Shohamy and Adcock, 2010; Wimmer et al., 2012).

Choice and memory

Behaviorally, we found that memory for incidental pictures was strongly predicted by choice selection, such that an object presented in a chosen option was much more likely to be recognized later than an object presented in a non-chosen option. The increase in memory for choices may be related to value: participants’ goal was to choose the option that they currently thought was most likely to lead to reward, and thus at this level choice is a proxy for valuation. While choice predicted memory, choice values estimated from a reinforcement learning model did not significantly predict memory; this null effect may be due to limitations in the reward learning model or in model fitting.

In the VMPFC, a region commonly found to correlate with value (Daw et al., 2006; Plassmann et al., 2007; Hare et al., 2008), we found an overlapping neural correlate of memory for chosen items and choice value. Interestingly, we found that the neural
correlate of choice memory did not relate to correlates of choice value: activity during remembered and forgotten choices was similarly correlated with choice value. Future studies targeted at memory and value may be necessary to better understand these signals in the VMPFC.

In conjunction with value, the benefit of choice on memory may also be related to attention. During decision making, it has been demonstrated that participants are biased to look more at chosen versus non-chosen options (Krajbich et al., 2010). In the current experiment, we did not record eye-tracking data, but it is likely that participants exhibited differential looking toward the chosen option, which may have contributed to enhanced memory. Interestingly, the effect of visual fixation during decision making has been correlated with activation in the VMPFC (Lim et al., 2011).

Episodic memory for chosen options was also predicted by activation in the hippocampus, a region critical for relational memory formation (Davachi, 2006). Activation in these regions of the hippocampus did not predict successful memory formation for non-chosen items. The differential effect of hippocampal activation on memory for chosen vs. non-chosen items suggests that the hippocampus does not blindly encode experiences: hippocampal responses are modulated by choice and likely by active behavior. This interpretation is supported by a recent study that demonstrated greater memory for items experienced during volitional exploration and intentional encoding (Voss et al., 2011), an effect related to enhanced hippocampal-cortical functional connectivity. Our results further demonstrate that trial-by-trial choice engage encoding processes in the hippocampus.
**Reward and memory**

During reward learning, surprisingly, we found that reward feedback had a negative influence on episodic memory. For high values of prediction error, large positive prediction errors decreased the likelihood of successful memory while large negative prediction errors increased the likelihood of successful memory. Overall, reward prediction error negatively but non-significantly modulated memory. The negative effect of reward result is counter to a hypothesized positive effect of reward on memory suggested by related studies (Wittmann et al., 2005). However, the negative effect of reward replicates a behavioral finding of a significant negative effect of reward on choice stimulus memory in a similar experiment (Chapter 4). The current finding is weaker; this may be due to the relatively low rate of overall memory, possibly caused by the change in environment from the scanned encoding phase to the non-scanned probe.

Neurally, we found that episodic memory formation modulated a reward prediction error signal in the striatum. In the ventral putamen, a region correlated with prediction error overall, we found that prediction error was only expressed for unsuccessful memory trials (**Figure 5.4**). Conversely, in a more anterior region of the striatum in the caudate, we found a positive interaction of reward and memory (**Figure 5.5**); here, prediction error was only expressed when stimuli were successfully encoded. The same region predicted individual differences of the influence of reward on memory: participants with a reward enhancement of memory exhibited differential reward prediction error signaling for remembered vs. forgotten stimuli (**Figure 5.6**). Reward prediction error-related activation in human fMRI has been demonstrated in many different types of reward tasks (O'Doherty et al., 2007). However, to our knowledge, the
modulation of reward prediction error by memory encoding processes has not been shown.

Based on prior studies of memory, we predicted that reward would have a positive effect on episodic memory. Prior fMRI research has found that incidental stimuli that signal high versus low potential reward value show increased memory and greater activity in the hippocampus and midbrain (Wittmann et al., 2005). Also, in an intentional encoding paradigm, stimuli presented on high value trials were more likely to be remembered (Adcock et al., 2006), an effect predicted by correlated activity in the hippocampus and midbrain.

Our finding of a negative effect of reward on memory is also surprising given the role of dopamine in memory formation in the hippocampus (Shohamy and Adcock, 2010). Dopamine has been shown to be critical for the maintenance of long-term neural connections in the hippocampus (Huang and Kandel, 1995; Otmakhova and Lisman, 1996; Bethus et al., 2010). Phasic fluctuations in dopamine in the striatum have been related to reward prediction errors (Schultz, 2006). However, the timescale of dopamine fluctuations in the hippocampus has not been resolved and thus it is difficult to make strong predictions about the effect of dopamine on memory in the hippocampus (Shohamy and Adcock, 2010). Future studies may reveal the mechanisms underlying the effect of reward on memory by investigating the effect of memory probe delays and the effect of presenting incidental stimuli at choice or at feedback.

During reward learning, how does memory formation interact with reward prediction errors? Ongoing neural processes responsible for memory encoding may influence striatal activity coding reward prediction errors. By the time of reward
feedback, memory encoding processes are already underway (as evidenced by activation predicting memory in the hippocampus and posterior cortex, Figures S5.1-S5.2). These memory processes may interact with striatal coding of prediction errors via cortico-striatal and hippocampal-striatal connections (Haber and Knutson, 2010; Pennartz et al., 2011). Alternatively, reward prediction error signaling may promote or inhibit neural systems responsible for memory encoding. While we did not find correlates of connectivity changes between the striatum and hippocampus during memory formation, functional connectivity could provide further information about the interaction between memory processes and reward prediction error. Connectivity may also illuminate the different patterns of memory and prediction error interactions we observed in the putamen and caudate.

**Conclusion and future directions**

In future work, it will be of interest to probe whether and how episodic memory formed during reward learning influences later decision making. It is possible that the fleeting experience of reward and choice during learning leaves traces in memory. Supporting this idea, a recent fMRI study found significant traces of value from a reward-motivated encoding task at re-exposure (Kuhl et al., 2010). In a future encounter, it is an interesting question whether remembered items could push or pull behavior based on traces of choice and value from initial encoding. For example, if remembered choices reactivate the VMPFC region associated with successful encoding, could this activity shift valuation processes in the VMPFC and bias decision making?
In conclusion, we find that episodic memory formation is modulated by ongoing reward learning. The idea that memory, reward, and decision making are tightly related is consistent with recent research that has begun to show how episodic memory guides future behavior (Buckner, 2010). A strength of the current approach is that episodic memories were formed during ongoing, goal-directed reward learning. This approach allows us to concurrently probe both reward learning and episodic memory. Our results demonstrate that during reward learning, the brain encodes more than just choice values and feedback. Behaviorally, episodic memory was increased for chosen options. Surprisingly, memory was negatively modulated by reward feedback. Neurally, we find that regions coding value and reward are also modulated by successful memory formation. The integrative study of memory formation and learning from feedback promises to provide further insight into the function of different memory systems in the brain.
Figure S5.1. Hippocampus and memory.

Activation in the left hippocampus during choice predictive of overall subsequent memory ($z = 3.97 (-26, -26, -12)$; $P < 0.05$ SVC).

Figure S5.2. Occipital and parietal cortex and memory.

Activation in the occipital and parietal cortex during choice predictive of overall subsequent memory (right, $z = 4.15 (40, -88, -2)$, left, $z = 4.22 (-38, -84, 12)$). ($P < 0.05$ SVC, ROI derived from contrast of task with objects vs. task without objects.)
Figure S5.3. Choice and feedback activation.

Activation correlated with choice value, A, and reward prediction error, B, during the reward learning task. (P < 0.005, uncorrected; reward prediction error masked for the striatum.)
Chapter 6: General discussion
Discussion

In the studies presented in this dissertation, I have demonstrated novel interactions between learning and memory systems during learning and decision making. These studies further our broader understanding of learning systems, how they interact, and how they guide behavior. The studies described in Chapters 2 and 3 demonstrated that relational representations, supported by the hippocampus, interact with the striatum and affect reward learning and value-based decision making. The studies reported in Chapters 4 and 5 demonstrated that the formation of long-term episodic memories is affected by processes previously thought to be the exclusive domain of incremental stimulus-reward learning, including choice, value and reward feedback.

In Chapter 2, I investigated the neural systems supporting generalization of value information across options in an environment where stimuli are related by similar reward likelihood (Wimmer, Daw, and Shohamy, 2012). The striatum, as shown previously, contributed to reward learning. Interestingly, however, striatal signals related to reward learning also reflected generalization across options, based information beyond each option’s reward history. Activity in the hippocampus reflected a different aspect of learning, coding choice values. Functional connectivity between the striatum and hippocampus predicted the degree to which participants generalized reward across options. Thus, although striatum and hippocampus are traditionally associated with reward learning and relational learning, respectively, I observed reflections of both sorts of learning in both systems.

Next, in Chapter 3, I explored the interaction of relational and reward learning systems in to further probe how relational representations guide decision making
(Wimmer and Shohamy, *in press*). This experiment was designed to test the hypothesis that relational memory processes in the hippocampus can support shifts in value and bias later decision making. Here, I found that activation in the hippocampus during reward learning indeed predicted a shift in value for previously associated stimuli. Additionally, this behavior was predicted by reactivation of associations in specific visual cortical regions as well as by functional connectivity between the hippocampus and striatum.

Finally, in Chapter 4 and 5 I explored how choice and reward learning influence memory formation. Behaviorally, these studies demonstrated that memory formation during reward learning was significantly modulated by reward, as well as choice. Intriguingly, a 24-hour delay between learning and the memory test revealed changes in the strength and direction of the influence of reward on memory. For pictures presented during choice, reward enhanced memory at the immediate test but impaired memory at the delayed test; for pictures presented during feedback, reward enhanced memory both immediately and after a delay. Chapter 5 explored the neural correlates of the effect of choice and reward on episodic memory. Neurally, the enhancement of choice picture memory by choice was related to activation in the VMPFC and hippocampus, while the surprising impairment of memory by reward was related to a modulation of prediction error activity in the striatum.

**Limitations and Future Directions**

*Limitations*

The research studies presented in this dissertation have several limitations. First, the results on the neural substrates of learning and memory are based on fMRI data. The
fMRI BOLD signal provides correlational evidence for neural involvement in learning and memory, but data from this method cannot show that a given area is necessary for relational or reward learning. However, it is important to note that our interpretation of these results is based upon and converges with evidence from lesion and electrophysiology studies in animals. For example, in the case of decision bias in Chapter 3, my results on the role of the hippocampus in learning are supported by prior animal lesion studies that show a necessary role for the hippocampus in similar behaviors (e.g. Port and Patterson, 1984; Port et al., 1987). Importantly, fMRI can demonstrate precisely when learning systems are engaged, evidence that these lesion studies cannot provide.

Additionally, my fMRI results that suggest an interaction between learning systems only provide correlational evidence for an interaction. For example, the finding in Chapter 3 that functional connectivity between the hippocampus and striatum predicts decision bias cannot demonstrate that the correlation between the hippocampus and striatum plays a causal role in driving decision bias. Future research studies using different methodologies such as patient studies will be necessary to determine the causal role of the hippocampus and striatum in these processes. Notably, the studies in Chapter 2 and 3 make use of paradigms that can be easily tested in animals. Thus, future electrophysiological or optogenetic studies in animals may provide an important test of the necessary role of different brain regions for these behaviors, complementing the fMRI results in humans.

A further limitation is that these studies were all conducted using controlled experimental paradigms and environments. It will be of interest to explore how the findings presented in this dissertation connect to real-world behavior. A primary
motivation of these studies was to investigate learning in situations more similar to everyday experience than many prior studies of learning and memory. In my experiments, both reward and relational information is often concurrently available. Even though the results suggest that different learning systems are indeed engaged in decision making, it is possible that the experimental situations do not generalize to how decision making operates in real-world situations. Further, the experimental results themselves may have limited generalizability to the larger population, as our studies were conducted on a young adult group largely drawn from a selective university community.

Finally, this research may be limited as they rely on using monetary rewards as positive feedback. It will be of interest to explore whether these results extend to learning and decision making when the rewards used are not only monetary gains. While monetary rewards have been shown to elicit similar patterns of brain activity as desired products, foods, and primary rewards such as juice (e.g. Chib et al., 2009), the implications of our results for decision making would be more generalizable if other forms of reward were used. Similarly, it will be important to verify whether these results extend to aversive feedback such as monetary loss. Of interest, prior results in the decision bias paradigm hint that monetary loss leads to a weaker decision bias (Wimmer and Shohamy, 2011).

In particular, regarding the effect of reward learning on memory in Chapters 4 and 5, it is not known if similar behavioral and neural effects would be found in a loss-avoiding context instead of a reward-seeking context. Future studies could determine if loss has the same impact on memory as a missed reward, and loss avoidance the same impact on memory as reward. Loss avoidance has been shown to elicit similar activation
as reward in the VMPFC (e.g., Kim et al., 2006), which may indicate that similar choice and reward effects on memory would be found for monetary losses. However, loss avoidance motivation has recently been shown to engage a distinct neural circuit from reward motivation during motivated memory encoding (Murty et al., 2012).

**Future Directions**

In future research, I plan to further explore the role of different learning systems in decision making, and their potential interactions, in several different directions.

*Effects of age and disease on relational and reward learning*

The interaction of relational learning and reward learning, as demonstrated in the decision bias paradigm in Chapter 3, may be impacted by age, disease, and addiction. In ongoing work with collaborators, I am investigating the role of dopamine and aging in the decision bias paradigm using a Parkinson’s disease (PD) model of dopamine depletion. Preliminary results show that PD patients have intact reward learning as well as an intact shift in value. Older adults, in contrast, do not show a significant shift in value. This result suggests that decision bias in PD patients may depend different cognitive and neural systems.

In future work, it will be of interest to explore the connection between the use of relational learning in decision making and disorders that impact decision making such as addiction. Over habitual use, stimuli and environments linked to addictive drugs become highly valued (e.g. Redish, 2004). In a process similar to the proposed mechanism for decision bias, associative memory processes in the hippocampus could lead to rapid shifts
in value of stimuli in the network of experiences connected to drug use. Such an interaction of associative memory processes and strong reinforcers may lead to systematic changes in decision making, and a greater understanding of these processes may suggest novel routes for treatment.

*Decision making and consolidation of learning*

In the future, it will be essential to study how the representations formed by relational as well as reward learning systems change over time. The consolidation processes that lead to long-term memory maintenance may begin with replay of learning events during rest and sleep (Tambini et al., 2010; Carr et al., 2011). And as noted in Chapter 1, consolidation over a short time-span of hours and days may be facilitated by dopamine in the hippocampus. Here, dopamine has been shown to facilitate cellular traces of memory after hours but not immediately (Huang and Kandel, 1995; Otmakhova and Lisman, 1996; Bethus et al., 2010). Over the long-term, different neural systems may be engaged in the maintenance of representations in memory. Representations encoded rapidly by the hippocampus may be slowly shifted to storage in other cortical regions (e.g., McClelland et al., 1995; Marshall and Born, 2007).

For decision making, memory consolidation may particularly have an effect on the interaction between relational and reward learning described in Chapter 3. Using a decision bias task similar to the paradigm used in Chapter 3, I am currently exploring the effect of a delay between learning and the probe of decision bias. These studies will probe whether consolidation of relational associations, or relational and reward associations, modulates the shift in value.
Memory for reward associations and decision episodes

Building on the studies described in Chapter 4 and 5, I am interested in further exploring the effect of reward learning on episodic memory, as well as the effect of episodic memory on later decision making. In particular, I am interested in whether people are able to recollect the experience of reward or miss feedback associated with the fleeting experience of items during learning. Further, I am interested in the broader question of whether and how people remember prior decision episodes.

In a follow-up study of the effect of reward learning on episodic memory, I am currently testing whether participants exhibit significant memory for reward and miss feedback associated with trial-unique pictures. Even though pictures are only seen once, initial data from a reward learning task with pictures presented during feedback suggests that reward and miss associations are remembered at above-chance levels. Significant memory for reward and miss associations would indicate that the impact of reward learning on episodic memory is even more extensive than initially found.

Memory for and neural traces of reward and miss feedback may also be explored in future fMRI studies. For example, with fMRI data collected during a memory recollection probe after reward learning, analyses could detect neural correlates of reward feedback associations. Supporting the existence of memory for reward associations, in a study where items were presented for motivated encoding, later re-presentation elicited traces of the initial value associated with the item (Kuhl et al., 2010). A demonstration of memory encoding and recollection of value would have important implications for later decisions that could be based on this value memory.
Separately, I am interested in how people encode memory for decisions. In particular, I am interested in memory for the choice situation, options, and estimated values of options – what could be called a “decision environment”. For example, consider the experience of making a decision between multiple important options in life. Each is associated different future actions and estimates of success. After choosing one option and receiving feedback, what remains in memory of the original decision environment? Can this kind of memory support updating of option values and later decision making, similar to model-based reinforcement learning? How do neural systems of relational and reward learning support this kind of memory? Understanding how the brain stores representations of decisions and values from past experience is an important area of future research.
References

Materials and methods are available as supplementary material on Science Online.


Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Netw 12, 961-974.


Granado, N., Ortiz, O., Suarez, L.M., Martin, E.D., Cena, V., Solis, J.M., and Moratalla, R. (2008). D1 but not D5 dopamine receptors are critical for LTP, spatial
learning, and LTP-Induced arc and zif268 expression in the hippocampus. Cereb Cortex 18, 1-12.


McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102, 419-457.


