THE NEURAL SUBSTRATES OF DETERMINISTIC DECISION-MAKING

by

Joshua J. Tremel

B.S., University of Pittsburgh, 2009

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Joshua J. Tremel

It was defended on

July 25, 2018

and approved by

Marc Coutanche, Assistant Professor, Psychology

Brent Doiron, Professor, Mathematics

Timothy Nokes-Malach, Associate Professor, Psychology

Dissertation Advisor: Julie Fiez, Professor, Psychology

The Neural Substrates of Deterministic Decision-making

Joshua J. Tremel, PhD

University of Pittsburgh, 2018

When making a decision, we draw upon multiple mnemonic resources to inform our behavior and to ideally produce a good outcome. Multiple memory systems guide this process, including a medial temporal lobe (MTL) system and a striatal system. The MTL provides episodic details about specific instances of prior experience, whereas the striatum provides a prediction about possible outcomes based upon a fusion of many prior experiences. While both of these systems are assumed to support decision behavior, extricating their discrete contributions has been challenging. Using neuroimaging and computational reinforcement learning, this study investigated the extent to which the MTL and striatal systems are co-active during single-exposure learning and how these systems each support subsequent behavior. This was done in the context of a single-exposure deterministic decision-making task that separated encoding processes from subsequent decision-making processes. Human subjects learned to associate words with monetary feedback in a single decision experience. They then used that information to make better choices in a subsequent round without feedback. Activity in MTL regions predicted episodic memory accuracy and correlated with subsequent decision accuracy and response times. Additionally, the MTL supported a model-based reinforcement learning process wherein initial decision experiences were used to build a model of the environment that was then used to prospectively formulate future decision outcome predictions. Activity in striatal regions also correlated with subsequent decision accuracy and response times, but did not relate to memory accuracy. The striatum supported a model-free reinforcement learning process wherein

predictions about decision outcomes were generated from a retrospective accumulation of prior decision experiences. Together, these results implicate both the MTL and striatum as essential substrates to single-exposure learning, but underscore that these systems operate in fundamentally different ways. The MTL is associated with prospective learning, wherein single instances of prior experience can be leveraged to inform subsequent choice. The striatum, in contrast, is associated with retrospective learning, wherein a history of experience is required to build reliable predictions about subsequent choices. In combination, the MTL system seems to support decision behavior until the striatal system has had enough experience to refine predictions about outcomes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1.0 INTRODUCTION

Our world is filled with stimulus-response-outcome relationships that we learn about and use to guide our behavior. While some of these relationships are probabilistic, many are fully or effectively deterministic. This dissertation is a part of a program of research investigating the neural basis of deterministic decision-making. In prior work, I have found that deterministic decisions can be guided by different types of information deriving from separate neural systems. On one hand, a striatal reinforcement learning system can generate and update predictions about choice outcomes based on a history of relevant experiences (Daw & Doya, 2006; Dayan & Daw, 2008; Ito & Doya, 2018; Jocham et al, 2011; Niv, 2009; O'Doherty et al., 2017; Packard & Knowlton, 2002). Learning via this system is thought to be retrospective, wherein prior experience is represented as an accrued history of prior outcomes in a composite score of the value of a choice. While this composite value is based on instances of experience, only the value, and not the individual instance, is retrievable. On the other hand, a medial temporal lobe (MTL) system can encode details about individual decision experiences and associate those details with particular outcomes. In contrast to the striatal system, learning via the MTL system is thought to be prospective, wherein individual instances of prior experience can be retrieved to predict future outcomes and guide decision-making accordingly (Davachi & Wagner, 2002; O'Doherty et al., 2017; Squire, 1992; Squire & Zola, 1996; Squire & Zola-Morgan, 1991). While it is assumed that deterministic decision-making involves contributions from both of these systems (Delgado

& Dickerson, 2012; Dickerson & Delgado, 2011), how and whether these systems work together to optimize deterministic choices in practice is unknown.

In the first study of this research program (Tremel et al., 2016), I examined the overlap of neural substrates underlying learning and decision-making in deterministic settings. The goal was to better understand the mechanism by which choice experiences translate into improved decision-making. Using a combination of computational modeling and functional neuroimaging (fMRI), this study was able to link sub-regions of the striatum to different learning, memory, and decision processes. Specifically, the dorsolateral striatum (putamen) acted as a mnemonic storage site representing predictions about choice outcomes formed from repeated decision experiences. In contrast, the dorsomedial striatum (head of the caudate nucleus) acted as a performance monitor, responding to outcome prediction errors and signaling for adjustments to decision control processes. However, the MTL system was conspicuously absent from these contrasts, suggesting perhaps that the computational approach was incapable of modeling MTL contributions, that MTL contributions were overshadowed by striatal contributions in this particular task, or that the MTL was less involved in deterministic decision-making than initially thought.

Nevertheless, neuropsychological evidence specifically implicates the MTL in deterministic decision-making. Damage to the MTL routinely impairs performance on deterministic decision-making tasks, such as concurrent discrimination learning (Buffalo et al., 1999; Corkin, 2002; Hood et al., 1999; Squire et al., 1988; Squire and Zola, 1996; Zola-Morgan et al., 1989; Zola-Morgan et al., 1994). Though learning can be preserved in some cases of MTL damage, successful performance normally requires a long trajectory of repeated experiences (Bayley et al., 2005; Buffalo et al., 1998; Chudasama et al., 2008; Gaffan & Murray, 1992; Hood

et al., 1999; Malamut et al., 1984; Phillips et al., 1988; Suzuki et al., 1993). Given this, it is possible that the role of the MTL in deterministic decision-making is to leverage individual instances of experience early in the learning trajectory to build an initial memory foundation to support later learning. Without this foundation, like in cases of MTL damage, learning may take longer by having to rely on retrospective habit learning via the striatum (Bayley et al., 2005) or other types of gradual learning processes associated with the acquisition of semantic information (Duff et al., 2006; Kan et al., 2009; O'Kane et al., 2004; Sharon et al., 2011).

Thus, given this evidence, it seemed that the best method to investigate MTL contributions was to specifically manipulate the efficacy of MTL-based (i.e., declarative) memory. The second study in this program explored the question of how declarative memory, via an MTL system, helped to improve subsequent deterministic decisions. In this study, I examined the same deterministic learning task as in the first study (concurrent discrimination), but included a list-length manipulation, wherein one group of subjects learned a set of 50 items over eight repetitions and another learned a set of 100 items. The hypothesis was that subjects learning a smaller set of items would have superior declarative memory compared to those learning the larger set of items (Jacoby, 1991; Mahut et al., 1982; Mishkin, 1982; Wais et al., 2006; Yonelinas, 2001), and therefore would place a greater emphasis on utilizing an MTL-based approach to guide decision-making on this task. Both groups of subjects exhibited behavioral profiles of memory consistent with the hypothesis, and the group learning the larger item list exhibited greater striatal activity than the shorter list group. This shorter list group exhibited learning-related activation in an MTL region in the hippocampus proper, even though this MTL activity was unassociated with computationally derived measures of the decision-making process. Thus, this study provided evidence that declarative memory and the neural

systems underlying it may contribute to deterministic learning, but questions remained with respect to how these mnemonic processes, specifically those mediated by the MTL, fundamentally relate to subsequent choice making episodes. Moreover, there was still a lack of clarity regarding the apparent discrepancies between these imaging studies and neuropsychological findings that point to the MTL as a necessary substrate for deterministic decision-making.

Importantly, concurrent discrimination is an extended repetitive learning task, wherein feedback is delivered after every decision experience. This encourages the re-encoding and updating of memory during each decision event. Thus, processes related to learning via repetitive feedback (i.e., habit learning) may underlie successful concurrent discrimination learning. This may be why these two prior studies found a strong relationship between striatal activity and behavioral and computational measures of decision-making. As such, it is possible that the role of the MTL might be masked by the long, repetitive learning trajectory of concurrent discrimination. It may be the case that the MTL is necessary early in this trajectory to set up an initial memory scaffold that can be used to produce early decision successes. As learning continues, the slower striatal procedural memory system can take over after enough experience (e.g., a few repetitions of correct outcomes). Once the striatal system takes over, the MTL scaffold may no longer be necessary.

The overarching goal of this dissertation is to bridge this apparent disconnect between the neuropsychological literature and previous computational neuroimaging findings (Tricomi & Fiez, 2008, 2012; Tremel et al., 2016, 2018). This dissertation investigates the possibility that the MTL and striatum operate in parallel and are co-active early in the learning trajectory of deterministic decision-making (i.e., after a single exposure to a decision outcome). To

accomplish this, I present an empirical study of single-exposure deterministic decision-making that simultaneously manipulates the engagement of the MTL and striatal systems in a within-subjects design. Subjects performed a task in an fMRI scanner, in which they learned to associate words with positive or negative feedback in a set of 80 word pairs. In the first exposure to these decision experiences (Round 1), subjects guessed at which word was correct and received feedback about their decision. In a subsequent round (Round 2), they then used this prior experience to guide their decision-making and make better choices. Decision-making in the second round is presented in the same context as in the exposure round, but without feedback, preventing subjects from learning and updating via new information. Thus, this design separates encoding and learning processes in Round 1 from decision-making processes in Round 2. This task also featured a 2 x 2 factorial manipulation of monetary reward magnitude and associative context to differentially target memory systems in the striatum and MTL, respectively.

The first chapter (Chapter 2) dissociates MTL and striatal contributions and investigates the system-level interactions during deterministic decision-making. In this chapter, I analyze the effects to two experimental factors – associative context (via images) and reinforcement magnitude – that have the intended purpose of modulating the relative engagement of the MTL and striatal systems. Activity in the MTL was expected to predict the accuracy of subsequent decisions and to be associated with episodic memories of specific decision experiences, reflecting that the MTL supports the encoding and retrieval of event-specific details that can inform later choices. Activity in the striatum was expected to predict the accuracy and speed of subsequent decisions and to be associated with familiarity-based memories of specific items, reflecting that the striatum supports the accumulation of an item-specific reinforcement history that can contribute to decision response automaticity. The goal of this chapter was to establish

that the MTL contributes to deterministic learning alongside the striatum and that these MTL contributions are especially important in single-exposure deterministic decisions.

The subsequent chapter (Chapter 3) tests the computational mechanisms associated with the MTL and striatal systems and whether the MTL might build an initial memory scaffold to support initial learning and decision-making. To do this, I simulated different types of reinforcement learning agents using data from Chapter 2. One type of reinforcement learning agent, model-free, implements a retrospective process wherein instances of decision outcomes are accrued through repetition into a metric representing the overall predicted value of a choice. In this type of learning, individual experiences are not directly used to guide decision-making, but rather used to build a history of observed reinforcement associated with a particular choice. In contrast, model-based reinforcement learning implements a prospective learning process wherein initial experience is used to build a model of the decision environment which is subsequently used to predict outcomes of future choices. If the above distinction between striatum- and MTL-mediated contributions to decision-making is correct, activity in these regions should map onto the computational distinction between model-free and model-based reinforcement learning, respectively. Furthermore, if the MTL is the primary source of knowledge early in the learning trajectory and therefore essential to initial decision-making, subjects who perform well in the decision task and subsequent memory tasks should exhibit a greater preference for model-based learning relative to poor performers. Taken together, the third chapter examines the possibility that the MTL is vital to successful decision-making after just a single prior choice experience and that the striatum is associated with making decision responses more habitual with repeated experience.

2.0 MEDIAL TEMPORAL LOBE AND STRIATAL MEMORY SYSTEMS UNDERLIE DETERMINISTIC DECISION-MAKING

Our environment is composed of pervasive decision-outcome relationships that we learn about through experience. Making a decision and experiencing its associated outcome provides essential information that can be used to optimize future choices. Several neural systems seem to underlie the encoding of these experiences, wherein different types of mnemonic representations are constructed to characterize different aspects of a particular decision event. Two systems in particular, the medial temporal lobe (MTL) associative memory system and the striatal procedural memory system, seem to play a role in translating prior experience into decision-making behavior, but extricating their discrete contributions has been challenging. This study manipulates two factors associated with each system in order to dissociate the contributions of the MTL and the striatum. This is done using a single-exposure deterministic decision-making task, wherein memory encoding and learning processes are separated from decision-making processes.

Prior work in this area has highlighted striatal involvement in deterministic decision-making, but the contributions of the MTL have remained elusive. Using a concurrent discrimination learning task, in which subjects learned through repetition to choose particular words associated with positive feedback, it was found that different sub-regions of the striatum implemented computational measures of learning (reward prediction errors and choice-value

predictions via reinforcement learning) and decision-making (decision thresholds and drift-rates via drift-diffusion modeling) (Tremel et al., 2016). Specifically, regions in the head of the caudate nucleus responded to errors in outcome value expectations (reward prediction errors) and were associated with performance monitoring (decision thresholds). In contrast, a region in the putamen was associated with bottom-up mnemonic evidence accumulation and storage (choice-value predictions and drift-rates). Taken together, it was hypothesized that these regions worked in concert to store and update underlying mnemonic representations that were then used to inform subsequent decisions (Seger & Cincotta, 2005; Seger et al., 2010). In contrast to the apparent involvement of striatal regions, MTL regions were unassociated with the measures of learning and decision-making considered in this study, leaving open the question of how the MTL contributes to choice performance.

A follow-up to this study used a list-length manipulation to modulate the efficacy of declarative memory and thereby impact contributions of the MTL (Tremel et al., 2018). List-length manipulations such as this have been traditionally used to reduce the efficacy of declarative memory encoding and retrieval (Jacoby, 1991; Mahut et al., 1982; Mishkin, 1982; Wais et al., 2006; Yonelinas, 2001). As such, two groups of subjects participated in a concurrent discrimination task with either 50 items or 100 items. It was expected that the 50-item group would outperform the 100-item group since they could theoretically rely on more robust MTL-based declarative memory retrieval to draw upon memories of individual instances of prior experience. However, while recognition and episodic memory performance indicated that the list-length manipulation succeeded in modulating the efficacy of declarative memory, activity in MTL regions was again unassociated with computational measures of decision-making (i.e,. decision thresholds from a drift-diffusion model), despite exhibiting a list-length effect.

Consistent with the prior study, activity in the caudate nucleus of the striatum seemed to play a key role in supporting decision-making for the longer list group, who exhibited poorer declarative memory performance overall. Thus, taken together, these two studies seem to suggest that the MTL may play a more limited role in deterministic decision-making compared to that of the striatum.

However, neuropsychological evidence indicates that the MTL not only contributes to these decisions, but in fact may be essential. Damage to the MTL routinely impairs performance on deterministic decision tasks, such as concurrent discrimination, in both monkeys and humans (Zola-Morgan et al., 1989; Zola-Morgan et al., 1994; Buffalo et al., 1999; Squire et al., 1988; Squire and Zola, 1996; Corkin, 2002; Hood et al., 1999). Though learning can be preserved in some cases of MTL damage, it often requires a longer trajectory of learning with more repetition to reach criterion levels of performance (Bayley et al., 2005; Buffalo et al., 1998; Chudasama et al., 2008; Gaffan & Murray, 1992; Hood et al., 1999; Malamut et al., 1984; Phillips et al., 1988; Suzuki et al., 1993). Thus, it is possible that the role of the MTL is to build an initial memory foundation that supports early decision-making after only a few exposures. Without this foundation, like in cases of MTL damage, learning may take longer by having to rely on gradual learning processes, such as a striatum-based procedural memory system (Bayley et al., 2005) or semantic memory systems (Duff et al., 2006; Kan et al., 2009; O'Kane et al., 2004; Sharon et al., 2011). When examining a longer trajectory of learning in healthy individuals, as in previous neuroimaging studies (Tremel et al., 2016, 2018), these putative early MTL contributions may be masked by consistent striatal engagement across the task.

To bridge the apparent disconnect between the neuropsychological literature and previous computational and neuroimaging findings (Tricomi & Fiez, 2008, 2012; Tremel et al., 2016,

2018), this study examines the possibility that the MTL and the striatum operate in parallel and are co-active early in the learning trajectory of deterministic decision-making. To do this, I simultaneously manipulate two factors that are expected to modulate the relative engagement of both systems. This is done in the context of a single-exposure deterministic decision-making task, designed to delineate mnemonic encoding processes from decision-making processes. This study, thus, can investigate the contributions of multiple memory systems after a single exposure without the drawbacks of examining repetitive learning across a longer learning trajectory. By targeting both systems with a factorial manipulation, different contexts of decision-making can be examined, wherein one system may be more effective than the other.

More specifically, to better understand the contributions of the MTL and striatum, subjects in this study performed this decision-making task during an fMRI session, in which they learned to associate words with positive or negative reinforcement in a set of 80 word pairs. Subjects learned these associations in a single exposure via feedback (Round 1) and used that experience to make better decisions in a second exposure (Round 2). This task also featured a 2 x 2 factorial manipulation of monetary reinforcement magnitude and associative context to differentially target memory systems in the striatum and MTL, respectively. Words in the task were associated with monetary reinforcement (reward or punishment) and with a detailed image to provide additional associative context. Detailed background images, such as an image of a natural landscape, have been used in other memory and decision-making tasks to enhance the engagement of associative memory mediated by the MTL (Bornstein et al., 2012; Bornstein et al., 2013; Doll et al., 2015b; Hannula et al., 2013; Hayes et al., 2010; Howard et al., 2011; Park et al., 2014). Likewise, monetary reinforcement has been preferentially linked to the engagement of striatum-based learning and memory systems (Jocham et al., 2011; Packard & Knowlton,

2002; Schönberg et al., 2007; Wimmer et al., 2012). These two factors were examined in the context of encoding, operationalized as the first round of the task (i.e., choosing an item the first time with no prior knowledge, but learning about its associated outcome), and subsequent decision-making, operationalized as the second round of the task (i.e., choosing an item a second time after just one episode of prior experience without additional feedback). As such, encoding and decision-making in this task represent different cognitive processes, with the former capturing the encoding of a current experience with the purpose of informing later behavior and the latter capturing the retrieval of prior experience to inform behavior.

The present study has two primary goals with respect to understanding the role of the MTL and the striatum in decision-making. First, this study seeks to establish links from activity in the MTL and striatum directly to decision-making behavior in the second round of the task using a regression approach with single-trial fMRI activity. It is expected that activity in both systems during subsequent decision-making (Round 2) will individually correlate to measures of decision behavior such as accuracy and response time. Statistically significant relationships between regional activity and decision behavior would implicate these regions as important contributors in supporting the retrieval and use of information to drive decision-making after a single experience. By directly correlating activity with decision behavior in Round 2 (i.e., after a single exposure), this approach avoids pitfalls associated with using computationally derived metrics of decision-making that can encapsulate additional non-decision processes such as additional encoding or executive roles.

As such, previous approaches have failed to distinguish between the instance-based retrieval of the MTL and the composite value history computed by the striatum. The second goal of this study is to make this distinction. This is done by examining encoding-related activity in

these regions during the initial decision experience in the first round of the task. The MTL is expected to contribute instance-specific information to decision-making and therefore should exhibit correlations between encoding-related activity and subsequent measures of episodic memory about individual items. In contrast, activity in the striatum is expected to not exhibit such a relationship since this system should rely on a composite history of prior outcomes and not instance-specific information. Thus, while both systems are expected to support decision behavior in the second round, each system should exhibit a unique memory profile of encoding in the first round, reflecting that different types of mnemonic representations are implemented by the MTL and striatum.

As a secondary aim, this study also seeks to characterize interactions between these systems and the factors that may modulate these interactions. I employ a psychophysiological interaction analysis testing the extent to which functional connectivity between MTL and striatal regions changes based on contextual factors in the task (i.e., associative context or monetary reinforcement magnitude). This analysis assesses whether activity in two given regions correlates in a manner that depends on these factors. A significant psychophysiological interaction would suggest that two regions exhibit shared processing that modulates based on a given factor (e.g., associative context). I expect the MTL to respond strongly to rich associative contexts, whereas the striatum should respond strongly to larger reinforcement magnitudes. If activity in MTL and striatal regions is correlated but depends on these factors, these systems may play parallel but distinct roles in supporting decision behavior. Altogether, this study provides evidence that the MTL and striatum have unique but fundamental roles in supporting deterministic decision-making behavior.

2.1 MATERIALS AND METHODS

2.1.1 Subjects

Twenty right-handed, native English speakers participated in a 2-hour behavioral and neuroimaging (fMRI) session. All subjects had normal or corrected-to-normal vision. Two subjects were excluded due to excessive movement in the fMRI scanner, leaving a sample of 18 subjects (10 female), who ranged in age from 20-25 years (M = 21.56, SD = 1.79). Subjects were compensated $60 for their time and could earn up to $15 in additional bonus money based on task performance. Informed consent was obtained from all subjects according to procedures approved by the University of Pittsburgh Institutional Review Board.

2.1.2 Deterministic decision-making task

2.1.2.1 Design

During the 2-hour experimental sessions, subjects first participated in a slow event-related, deterministic decision-making task in an fMRI scanner (Figure 1). In this task, subjects made decisions about 80 pairs of words and learned to associate particular words with positive or negative feedback (Round 1). They experienced these words a second time and used their prior experience to inform their decisions (Round 2). In Round 1, subjects selected one word in a presented pair and received feedback indicating whether choosing that item resulted in a monetary gain or loss. Round 1 outcomes for each pair were pre-determined to ensure an even split of gains and losses going into Round 2. In Round 2, subjects performed the same task as in Round 1, but used their knowledge about the items to choose the best word (i.e., the item

associated with a monetary gain). Pairs, however, were different in Round 2 such that whichever word was selected in Round 1 was presented in Round 2, but paired with a new word that had not yet been seen. In the case of a Round 1 gain, subjects would have to choose that item again to receive a gain in Round 2. In the case of a Round 1 loss, subjects would have to infer that the new word is the correct word and choose that instead, thereby avoiding a monetary loss. Round 2 items were re-paired in this way so that subsequent analysis could reduce the number of factors by focusing on selected items only, since selection can influence how decision experiences are encoded (Tremel et al., 2018). The deterministic nature of this task and the re-pairing in Round 2 were made clear to subjects during instructions and a short practice session.

Figure 1. Deterministic decision-making task structure.

In Round 1 (top panel) of the deterministic decision-making task, subjects were presented with a pair of words with an image in the background. After selecting one of the words with a key press, subjects received feedback indicated a monetary gain or loss associated with their choice. Images were either a clear natural landscape (displayed above) or a noise-degraded landscape image (associative context manipulation). Monetary reinforcement varied between low magnitudes ($0.10-$0.29) and high magnitudes ($0.81-$1.00). There was a fixed, equal ratio of gains and losses in Round 1. Round 2 (lower panel) began after subjects made decisions to each of the 80 pairs of words in Round 1. Only the word pairs were presented in Round 2, with no background image or feedback. Word pairs in Round 2 consisted of a word that was selected by the subject in Round 1 paired with a new word.

This task also featured a 2 x 2 factorial manipulation of monetary reinforcement magnitude and associative context to differentially target memory systems in the striatum and MTL, respectively. For the reinforcement manipulation, decisions were associated with a low or high magnitude reward or punishment, depending on accuracy of selection. Low magnitudes ranged from $0.10 to $0.29, while high magnitudes ranged from $0.81 to $1.00. Reinforcement magnitude was explicitly reported to subjects in the feedback they received in Round 1. For correct decisions, feedback displayed a plus sign and the magnitude in green text (e.g., "+$0.85"). For errors, feedback displayed a minus sign and the magnitude in red text (e.g., "-$0.85"). Subjects earned a monetary bonus on top of the base participation payment based on the cumulative value accrued in Round 2 performance. Round 1 was excluded from the bonus calculation since the outcomes are fixed and since subjects had no experience to draw upon at that point.

For the associative context manipulation, decisions were presented with a clear, detailed image or a noise-degraded image of a natural landscape. This image appeared in the background of Round 1 choice experiences and remained on the screen as feedback was delivered. Feedback and the context image were presented in Round 1, but not in Round 2. In Round 2, subjects only saw the words and made a selection, receiving no new feedback about their performance. Thus, Round 2 constituted decision experiences without new encoding of outcome information.

Of the 80 word pairs in Round 1, half were associated with the clear context and half with the noisy context. In each of these categories, half were then associated with high magnitude and half with low magnitude. Likewise, of those, half were predetermined to yield gains, while the other half were predetermined to yield losses. Words associated with gains or losses in Round 1

were associated with the same value and magnitude in Round 2 (though feedback was not delivered).

2.1.2.2 Word stimulus materials

Word stimuli (N = 640) were drawn from the MRC Psycholinguistic Database (Coltheart, 1981). All words were one-syllable, between three to six letters, and between two to five phonemes. Words were further constrained by four psycholinguistic characteristics: frequency (log HAL > 5.00, M = 9.10, SD 1.45), concreteness (range 350-700, M = 547, SD = 69.05), familiarity (range 350-700, M = 524, SD = 55.66), and imageability (range 350-700, M = 552, SD = 56.00). Additional word frequency data were acquired via the English Lexicon Project (Balota et al., 2007).

Words were divided into eight separate word lists of 80 words each. Each list was balanced for the above psycholinguistic criteria. To create the lists used in the deterministic decision-making task, two lists were randomly selected, randomized, and paired to create the word pairs. A third list was randomly selected, randomized, and used as the new word pairings of Round 2. The remaining five lists were set aside for use as distractor items in subsequent behavioral tasks. This routine ensured that each subject had a unique set of word pairs for the task, while remaining balanced for psycholinguistic characteristics.

2.1.2.3 Image stimulus materials

Images for the associative context manipulation (N = 80) were compiled from web searches (via Google Images) for public domain stock photos of natural landscape scenes. Images were processed to balance visual image properties across the set, such as contrast, luminance, and brightness. In the same processing routine, each image was also noise-degraded to produce a

duplicate set of noisy images for each clear image in the set. To balance visual image properties, the forward 2D discrete Fourier transform (DFT) of each image was computed using a fast Fourier transform algorithm (FFT). Each image's DFT was decomposed into a phase angle matrix and an amplitude matrix. The amplitude matrix was averaged across the entire set to normalize visual properties across images. For the clear image set, this average amplitude matrix was recombined with each image's individual phase angle matrix and inverted via inverse FFT to generate the final image.

For the noise-degraded image set, each original image was randomized to generate a random noise image with the same color (RGB) spectrum as the original image. Next, the phase angle matrix of this randomized noise image was recombined with the set's average amplitude matrix and inverted via inverse FFT to generate a fully noise-degraded image. This noise image was then combined with the corresponding processed clear image via linear interpolation at 75% percent noise (i.e., 75% noise image, 25% clear image). Figure 1 illustrates an example of a clear image and its noise-degraded counterpart. This procedure has been used for a similar purpose in studies of perceptual decision-making (Heekeren et al., 2004; Tremel & Wheeler, 2015).

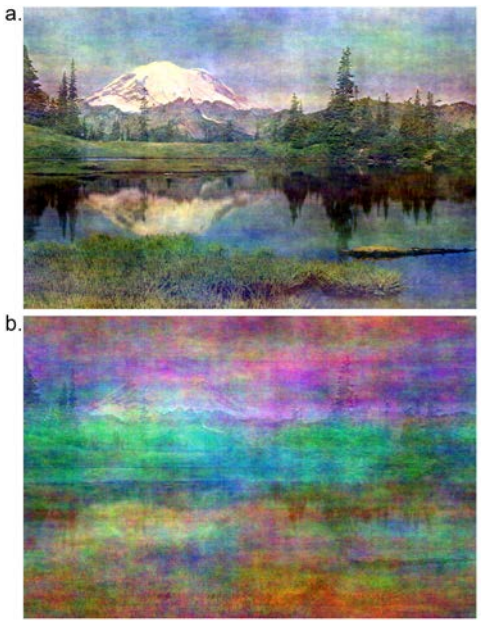

Figure 2. Associative context example.

(a) Example image of a clear associative context, consisting of a detailed image of a natural landscape. (b) Noise-degraded version of the same image. In the task, all presented clear and noise-degraded images were unique, such that a subject would never see the noise-degraded counterpart of a clear image used in the task.

2.1.2.4 Scan procedure

Before entering the fMRI scanner, subjects practiced a short, 8-pair version of the deterministic decision-making task and were instructed on the structure of the scan version of the task. Once entering the scanner, two sequences of anatomical images were acquired (T1- and T2-weighted), followed by functional imaging of the deterministic decision-making task. In the scan version of the task, Rounds 1 and 2 were broken up into four separate runs each of 20 trials (8 runs total). Each run lasted about 8.5 minutes.

For each trial, a pair of words was presented in a semi-transparent box in the center of the screen with a clear or noise-degraded natural landscape image in the background (Figure 1). Subjects chose one of the presented words with a button press during this 4 s decision response period. Button presses were assigned to the index finger of both hands, counterbalanced across

subjects. For half of the subjects, selecting the top word was assigned to a left-hand button press, while selecting the bottom word was assigned to a right-hand button press. This was reversed for the other half of subjects. After pressing the button, the chosen word was highlighted on the screen for the remainder of the 4 s decision epoch. This was followed immediately by a 2 s feedback period, in which feedback displayed choice accuracy (gain or loss) and associated monetary value in the box in the center of the screen. After feedback, subjects fixated on a cross at screen center for 2 s.

Because this task was a slow event-related design, a 14 s baseline period separated each trial. During this baseline period, it was possible that subjects could engage in a verbal rehearsal strategy instead of resting, thereby engaging the MTL and associated verbal working memory processes. This engagement could be detrimental to later analysis. Therefore, subjects performed a distractor 1-back task during this baseline period to prevent verbal rehearsal between trials. In this task, a series of six words, drawn from one of the unused word lists, were presented in sequence on the screen. A random word in a random place in the sequence appeared twice in succession. When this occurred, subjects were instructed to respond with a button press. Words were displayed on the screen for 1 s, separated by a 1 s fixation period.

In addition to the baseline period, trials were further separated by a variable inter-stimulus interval to facilitate group-level event-related analyses. This interval was sampled randomly from a positively skewed distribution ranging from 2-6 seconds, drawn in increments of 2 s (M = 3.2 s). Each run ended with a 12 s fixation period to allow the signal to fully decay to baseline after the final 1-back distractor sequence.

For the scan session, the task was projected onto a screen at the head of the magnet bore using a BrainLogics MRI Digital Projection System. Subjects viewed the screen via a mirror

attached to the radio frequency coil and indicated their response using fiber optic response gloves on either hand, connected to the presentation computer via a serial response box (BrainLogics, Psychology Software Tools, Pittsburgh, PA). Earplugs were provided to minimize discomfort from scanner noise. This and all other tasks were presented using the PsychoPy software package (Peirce, 2007; Peirce, 2009).

2.1.3 Functional localizer

At the end of the scan session, subjects participated in a functional localizer scan. This localizer was used to identify visual regions associated with stimulus-specific processing of scenes (parahippocampal gyrus), words (fusiform gyrus), and faces (fusiform gyrus). The goal of localizing these regions was to assess whether visual representations of items experienced in the task were related to processing in the MTL and striatum. Functional connectivity was computed to assess this by examining how functional coupling from regions of the MTL and striatum to stimulus-specific regions in the ventral stream changed as a function of factors in the task, such as associative context or reinforcement magnitude. Factor-dependent changes in functional connectivity between the target memory systems and visual regions would suggest that mnemonic processing in the MTL and striatum draw upon particular visual representations. For example, words associated with high value may be remembered better later, which might be associated with increased functional connectivity between the striatum and a word-selective visual region. Likewise, a clear landscape image may enhance associative memory, which might arise from increased functional coupling between the MTL and scene-selective visual regions. The face category was intended to localize a control region, since no face images were presented

in the task and were thus unassociated with task performance. Thus, face-selective regions should not be functionally related to processing in the MTL and striatum.

The localizer implemented a mixed-block design, wherein subjects viewed several stimuli of a single category in 16 s blocks followed by 12 s of passive fixation. There were four blocks for each category of faces, scenes, and words. Within each block, subjects performed a 1-back task similar to the one during the baseline period of the deterministic decision-making task. Seven images or words were displayed in each block, one of which repeated in sequence for the 1-back task. Each image or word was displayed for 1.5 s, separated by 0.5 seconds of fixation. Additional variable fixation separated the trials, ranging from 0-6 s (M = 2.5 s).

Word stimuli were selected from unused lists described above. Scene stimuli were a separate set from those described above, but acquired in the same manner. They were not processed for visual uniformity to ensure a clear and robust activation of place-selective regions. Face stimuli were a part of the MacBrain Face Stimulus set (courtesy of the MacArthur Foundation Research Network on Early Experience and Brain Development, Boston, MA).

2.1.4 Surprise memory test

After the scan session, subjects participated in a behavioral surprise memory test. In this test, subjects were presented with a single word at the center of the screen and responded to two memory probes. Words in this task were either words that were selected during Round 1 of the decision task (N = 80) or new words drawn from a list of unused words (N = 80). In the first memory probe (Recognition probe), subjects were asked whether they recognized the word from Round 1 of the decision task or not. Decision task words that were correctly recognized were considered "hits" (versus "misses"), while new words that were mistakenly reported as

recognized were considered “false alarms” (versus “correct rejections”). In the second memory probe (Episodic probe), subjects were asked to recall their experience from Round 1 and report whether they received positive or negative feedback when they encountered the presented word. This task was used to assess subsequent memory and the nature of the underlying memory representations that were used to inform decision-making.

Memory test results were analyzed behaviorally by categorizing recognition and episodic probe responses into three categories. The “no memory” category comprised items that subjects failed to correctly recognize. “Familiarity” reflected items that were correctly recognized (hits), but without episodic details. Last, the “episodic” category captured items that were correctly recognized (hits) with correctly recalled episodic details. Memory behavior was analyzed alongside imaging data described in Section 2.1.5.6.

2.1.5 Imaging analysis

2.1.5.1 Image acquisition

MR images were obtained using a Siemens Allegra 3-T system. Anatomical images were acquired using a T1-weighted MP-RAGE sequence (repetition time, TR = 1540 ms; echo time, TE = 3.04 ms, flip angle, FA = 8°; inversion time, TI = 800 ms; 1 mm^3 isotropic voxels, 192 sagittal slices) and using a T2-weighted spin-echo sequence (TR = 6000 ms, TE = 73 ms, FA = 150°, 0.78 mm^2 in-plane resolution, 38 axial slices spaced 3.2 mm apart). Functional images sensitive to the BOLD contrast were acquired using a whole-brain echo-planar T2*-weighted sequence (TR = 2000 ms, TE = 25 ms, FA = 70°, 3.125 x 3.125 x 3.2 mm resolution, 38 slices spaced 3.2 mm apart). For each run, the first five images were discarded to allow system magnetization (three images) and radio frequency signal (two images) to reach a steady state.

2.1.5.2 Image preprocessing

Preprocessing and analysis of MR data was executed using FIDL (Washington University in St. Louis). The preprocessing routine adjusted functional images to account for noise and image artifacts. This included adjustment of within-TR slice-time acquisition to the temporal midpoint of the first slice via sinc interpolation, adjustment for motion using a rigid-body translation and rotation algorithm, intensity normalization within each run to a mode of 1000 to facilitate group-level comparisons, and transformation into Talairach atlas space with resampling into 2 mm^3 isotropic voxels (Fox et al., 2005; Lancaster et al., 1995; Ojemann et al., 1997; Snyder, 1996). The motion adjustment, normalization, and atlas transformation were computed and applied at the same time to avoid resampling the data more than once. All analyses were carried out in Talairach atlas space (Talairach and Tournoux, 1988).

2.1.5.3 Region identification

To identify regions in the MTL and striatum related to deterministic decision-making, a secondary imaging dataset was examined that used a concurrent discrimination learning task (Tremel et al., 2016; Tremel et al., 2018). This task featured a similar deterministic reinforcement schedule, but involved learning relationships between word pairs and feedback across eight repetitions of the task instead of just one experience with feedback. This dataset was acquired from 33 subjects who learned non-monetary feedback associations to either 50- or 100-pairs of words. By using this dataset instead of the current one to define regions of interest, I sought to avoid circularity in the analysis stream (i.e., statistically defining regions and statistically testing those regions using the same data) and thereby define relatively unbiased regions to examine the current task. Because there were *a priori* hypotheses about the MTL and the striatum, region identification was limited anatomically to tissue encompassing these regions.

Using the first two rounds of the concurrent discrimination task from the Tremel et al. (2016, 2018) dataset, a repeated-measures ANOVA was computed to test for a main effect of time across all trials. This analysis examined modulations across an event-related average of 11 time points of a trial. A significant main effect of time indicated that BOLD activity, on average, changed over the course of a trial. In other words, this analysis revealed voxels that were active during the task, indiscriminate of function (i.e., it does not exclusively localize activity due to memory, learning, perception, etc.). A whole-brain image representing the main effect of time was then smoothed with a 2 mm full-width at half-maximum Gaussian kernel and corrected for multiple comparisons (minimum Z-transformed F-statistic of 4.0, $p < 0.05$, minimum of 12 contiguous voxels). Voxels whose activity exceeded a Z-transformed F-statistic of 7.5 were considered region peak coordinates, around which 8 mm spheres were grown. Voxels within these spheres that failed to pass the multiple comparisons correction (i.e., they did not appear in the corrected image) were dropped from the region spheres. The Talairach-space atlas coordinates of the center of mass for these corrected regions were then computed. Bilateral anatomical homologues were consolidated into single regions (e.g., two regions in the left and right head of the caudate nucleus were combined into one bilateral head of the caudate nucleus region).

2.1.5.4 Time series extraction

One advantage of the slow event-related design is that single-trial BOLD activity time series data can be extracted without having to deconvolve overlapping event signals. To extract activity time series data, a GLM was computed for each subject, in which signal drift was modeled by a linear term and baseline signal was captured by a constant term. No effects of interest were directly modeled in this GLM. Instead, trial effects plus noise were retained in the residual error of the

model. This residual time series was expressed as percent signal change from the baseline term for each run and extracted. The full experimental time series was then segmented into trial-level time series by joining 11 time points in sequence from the onset of each trial.

2.1.5.5 Functional role of the MTL and striatum in decision-making

The first goal of this study was to dissociate and establish the nature of the contributions of the MTL and the striatum to deterministic decision-making processes. In particular, activity in the MTL and striatum was used to predict decision behavior, measured as accuracy and response times (RT). To do this, two mixed-effects regression analyses were implemented for each identified region of interest. These statistical analyses and others throughout this study were conducted using the R environment for statistical computing (R Development Core Team, 2016), including the Companion to Applied Regression (CAR) (Fox & Weisberg, 2011), linear mixed effects modeling (lme4) (Bates et al., 2015), MuMin (Barton, 2009), and sjPlot (Lüdecke, 2017) packages.

In the first analysis, logistic regression was used to predict trial-level decision accuracy in Round 2 using regional BOLD activity from Round 2. The second analysis was similar, but used linear regression to predict trial-level decision response times (RT) in Round 2 using regional BOLD activity from Round 2. Subject was included as a random effect in each model (random intercept for each subject), while regional BOLD activity was modeled as a continuous fixed effect. The overall goal of these analyses was to examine how each region contributed to different behavioral metrics of decision-making, such as accuracy and RT, as predicted by trial-level fluctuations in the BOLD response. Thus, these analyses sought to establish particular decision-related roles for each of the regions of interest.

Importantly, these regression models were computed using Round 2 imaging data, wherein subjects exploited their prior experience to drive decision-making and were no longer encoding contextual details about the task (i.e., there were no background images or feedback in Round 2). Thus, activity in Round 2 reflected a relatively pure decision-making signal derived from prior experience, with minimal burden on further encoding. For each analysis, the dependent variable was computed as the mean activity at time points 4 and 5 of the time series. This mean value corresponded to the average time of peak activity during the decision and response epoch (i.e., trial), allowing for lag in the hemodynamic response. Therefore, this metric should adequately represent a given region's primary engagement during the decision period of this task (Ploran et al., 2007, 2011; Tremel & Wheeler, 2015). This predictor was mean-centered for these analyses.

Prior work has indicated that the mnemonic encoding of choice experiences is heavily influenced by feedback about the selected choice. A decision experienced in the context of positive feedback is encoded differently than one experienced in the context of negative feedback (Tremel et al., 2016; Tremel et al., 2018). Thus, Round 1 Outcome was included as a predictor in these regression analyses as a categorical fixed-effect (i.e., whether the decision in Round 1 resulted in a gain or loss). To account for individual variability in sensitivity to gains and losses, slopes for this factor were allowed to vary by subject. The inclusion of random slopes for the Outcome effect was statistically justified by comparing the model to an alternative regression model excluding random slopes.

It is worth noting that while the Round 1 Outcome factor was included in this analysis, other factors such as associative context and reinforcement magnitude were excluded. These additional factors were excluded for several reasons. First, these factors were hypothesized to

influence the engagement of the MTL and striatum primarily during encoding (i.e., Round 1), since the literature has largely examined these factors in that context (Bornstein et al., 2012; Bornstein et al., 2013; Doll et al., 2015b; Hannula et al., 2013; Hayes et al., 2010; Howard et al., 2011; Jocham et al., 2011; Packard & Knowlton, 2002; Park et al., 2014; Schönberg et al., 2007; Wimmer et al., 2012). It is thus unclear whether these expected differences during encoding would translate into behavioral differences during Round 2, especially given prior findings that modulating the efficacy of one system (e.g., the MTL) may not produce behavioral differences in decision-making (Tremel et al., 2018). Second, including too many factors would hinder the ability to detect a real relationship between physiology (i.e., BOLD signal) and behavior. Because there was clear evidence that prior outcome (i.e., gain or loss) could influence the direction of effects, the outcome factor was chosen. Last, all factors were individually examined in separate exploratory regression analyses, but including either or both of the associative context or reinforcement magnitude factors failed to improve the fit of the models and in several cases, reduced the quality of fit. In contrast including the factor of Round 1 outcome increased the fit quality, supporting the decision to include this factor in this analysis.

Outlier trials were identified based on the first time point of each full trial time series. If this baseline value was beyond the first or third quartiles by more than 1.5 times the interquartile range, that trial was considered an outlier and removed from analysis. Qualities of model fits were assessed for logistic models by computing concordance and for regression models by computing the conditional and marginal coefficients of determination (r^2). Concordance reflects when the predicted probability for the expected event or outcome is higher than that for the event or outcome not occurring, according to the model (Austin & Steyerberg, 2012; Harrell, 2015). The conditional r^2 for a mixed-effects regression reflects the proportion of variance attributable

to the fixed plus random effects in the model, while the marginal r^2 value reflects the proportion of variance attributable to the fixed effects alone. P-values were corrected for multiple comparisons across all models using false discovery rate (FDR) controlling procedures (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). Regions that failed to exhibit decision-related effects in this regression analysis were excluded from subsequent analyses.

2.1.5.6 Functional role of the MTL and striatum in memory encoding

By establishing a relationship between decision-making behavior and the activity of particular MTL and striatal regions, these regions can be linked to theoretical functional roles in a decision. However, understanding these roles in the context of a decision does not test how and what types of information these regions contribute to a decision. To understand how these regions encode different representations of prior experience, regional activity during Round 1 was examined and related to behavior on the post-scan surprise memory test. The goal of this analysis was to connect activity related to the encoding of Round 1 decision experiences to the subsequent quality of item-level memories. This connection would further characterize regions of the MTL and striatum in terms of the kinds of mnemonic representations supported and utilized by these regions.

To examine regional roles in memory encoding processes, an item-level analysis (item as a random factor) was implemented to test for a relationship between task factors (e.g., reinforcement magnitude and context) and Round 1 activity in each region. Item-level analysis was used to examine whether activity related broadly to particular memory profiles. For instance, if regional activity tended to be greater for hit items than for miss items, that region was likely to encode some representation of recognition memory. A subject-level analysis, however, may not detect that relationship due to individual differences in recognition memory ability. Because

memory encoding is a late, post-decision process, a different time window was selected for this analysis, corresponding to feedback delivery (mean of time points 7 + 8). This later epoch should capture post-decision processes including the encoding of episodic experiences and the binding of outcome information to stimulus information. Critically, these encoding processes occur after a selection had been made and after feedback has been delivered.

Two analyses were computed for each region: one to assess recognition memory and one to assess episodic memory. The dependent variable in each analysis was regional activity (mean of time points 7 + 8), while the independent variables were task-related factors such as associative context or reinforcement magnitude. Memory behavior, determined by behavioral responses on the surprise memory test, was also included as an independent variable in order to tie regional activity to particular memory categories. For recognition memory, memory accuracy was used as a factor, separating items into hits and misses, indicating items that were successfully recognized and those that were not, respectively. For episodic memory, memory response was used as a factor, separating items into whether subjects reported that they believed they had received a gain or loss during their Round 1 experiences.

For the analysis of the recognition memory probe, encoding activity in each region was tested with a 4-way ANOVA, including factors of memory accuracy (hit, miss), Round 1 outcome (gain, loss), reinforcement magnitude (high, low), and associative context (clear, noisy). The memory accuracy factor describes whether a subject correctly recognized an item from the deterministic decision-making task (hit) or failed to recognize it (miss). Separate ANOVAs were computed for each region of interest.

A similar ANOVA was computed for the episodic memory probe, testing encoding activity in each region against four factors, including memory response (subject responded

"gain," "loss" to the probe), Round 1 outcome (gain, loss), reinforcement magnitude (high, low), and associative context (clear, noisy). The memory response factor here described the behavioral response of a subject to the episodic memory probe and does not capture memory accuracy. Instead, this factor (i.e., whether subjects report that they experienced a "gain" or "loss") combined with the Round 1 outcome (i.e., whether subjects actually experienced a gain or loss in Round 1) factor describes memory accuracy. A region that encodes subsequent episodic memory accuracy, therefore, should exhibit a crossover interaction between these two factors.

Since the critical purpose of this analysis was to tie memory behavior to regional activity (i.e., memory activity), effects that failed to interact with memory accuracy/response factors were dropped from further interpretation. Thus, a main effect of reinforcement magnitude, for instance, would not inform whether activity relates to item-level memory behavior, but an interaction between reinforcement magnitude and memory accuracy would. As with the regression models, p-values were corrected for multiple comparisons across all models using FDR-controlling procedures.

2.1.5.7 Functional connectivity between regions

Having characterized ways in which regions in the MTL and striatum contribute to deterministic decision-making and what memory representations underlie those contributions, the third goal of this study was to probe the dynamic relationships among these regions. It is unclear whether these regions act in parallel to (i.e., independently), in support of, or in opposition to each other and how that interaction may change as factors in the task change. To assess this, a psychophysiological interaction (PPI) analysis was conducted using each of the MTL and striatum regions as seed regions. This analysis probed how the functional connectivity among

these regions changed as a function of the factorial manipulation of reinforcement magnitude and associative context.

One PPI analysis was completed using each MTL and striatal region as a seed region. Additionally, PPI analyses were computed separately to examine the effects of the reinforcement magnitude factor (high, low) and the associative context factor (clear, noisy). For each PPI analysis, three regressors were created. First, the seed regressor was defined as the full, mean-centered residual time series of a given region. Second, a task regressor was created by coding 6 s trial blocks at the onset of each trial. Each of these trial blocks reflected contrast coding for the task factor of interest. For the reinforcement magnitude factor, high magnitude trials were coded as +1 and low magnitude trials as -1. For the associative context factor, clear background trials were coded as +1 and noisy background trials as -1. This regressor was centered at zero. This task time series was then convolved with a hemodynamic response function to account for BOLD signal lag to temporally align it to the seed regressor time series. Last, a target region time series was defined as the residual time series of another region of interest (i.e., other MTL and striatal regions).

The PPI was modeled as the interaction between the task (psychological) and seed (physiological) regressors as they predict activity in the target region. Subject was treated as a random effect (random intercept for each subject). A statistically significant interaction between the task and seed regressors in this model would indicate that functional connectivity between the seed and target is modulated by reinforcement magnitude (or context, whichever is modeled in the task regressor). P-values were corrected for multiple comparisons across all models using FDR-controlling procedures.

Secondarily, functional connectivity was assessed for these target MTL and striatum regions to higher-order visual perception regions selective for particular stimuli (scenes and words, with faces as a control). The purpose of this secondary analysis was to gain additional leverage in describing the kinds of mnemonic representations encoded by the MTL and striatum and how those encoding processes change with factors in the task. To define stimulus-selective regions, the functional localizer data was analyzed at a group level using contrasts between the "scene" blocks and "face" blocks and between the "word" and "face" blocks. Bilateral regions in the parahippocampal gyrus were selective for scenes and a right lateralized region in the fusiform gyrus was selective for faces. These coordinates of these regions were in the vicinity of those defined in prior work (Tremel & Wheeler, 2015). These regions were used as additional target regions in the PPI analyses described above.

This localizer task analysis, however, failed to localize a candidate word-selective region in the fusiform gyrus. Reported coordinates from several studies from the visual word form area (VWFA) literature were averaged to create a candidate VWFA region of interest, centered at -43, -56, -10 (Cohen et al., 2000; Cohen et al., 2002; McCandliss et al., 2003). Using the main effect of time image from the concurrent discrimination learning dataset, a 10 mm sphere was drawn with these coordinates as the center. Voxels that failed to pass a multiple comparisons correction were dropped from the region. This resulted in a 155-voxel region in the left fusiform gyrus with a center of mass at -45, -57, -11 and peak activation at -43, -55, -10, consistent with the VWFA literature. This candidate VWFA region was used as an additional target region in the PPI analyses described above. It is worth noting that while the localizer analysis was unable to identify a statistically significant cluster using the "words" versus "faces" contrast, the regional

average activity in this putative VWFA region exhibited a significant preference for words over faces, $z_t = 2.09$, $p = 0.02$.

2.2 RESULTS

2.2.1 Behavioral description

2.2.1.1 Deterministic decision-making task

Subjects performed a two-round deterministic decision-making task in which they learned from experienced decisions in the first round to inform later choices in a second round. On average, subjects were able to learn to correctly select items in Round 2 at a rate of 64.2% (SD = 0.17). This was a statistically significant increase from the chance average of 49.89% (SD = 0.01) of Round 1, $t(17) = 3.65$, $p < 0.01$. Chance accuracy in Round 1 was not exactly 50% due to trials with missed responses. Response times (RT) also increased from Round 1 to Round 2 from 1.42 s (SD = 0.25) to 2.00 s (SD = 0.29), $t(17) = 8.14$, $p < 0.001$. This increase is consistent with prior findings indicating that retrieval of newly encoded information requires additional time relative to the encoding episode itself (Eichenbaum, 2001; Jacoby, 1991; Mandler 1980; Tremel et al., 2016, 2018). Figure 3 illustrates the factorial breakdown of Round 2 accuracy and RT based on Round 1 outcome (gain, loss), associative context (clear, noisy), and reinforcement magnitude (high, low). While these factors (e.g., outcome and magnitude feedback and context images) were not present in Round 2, Round 1 encoding experiences were expected to influence the relative contributions of MTL and striatal systems, which may translate into behavioral differences.

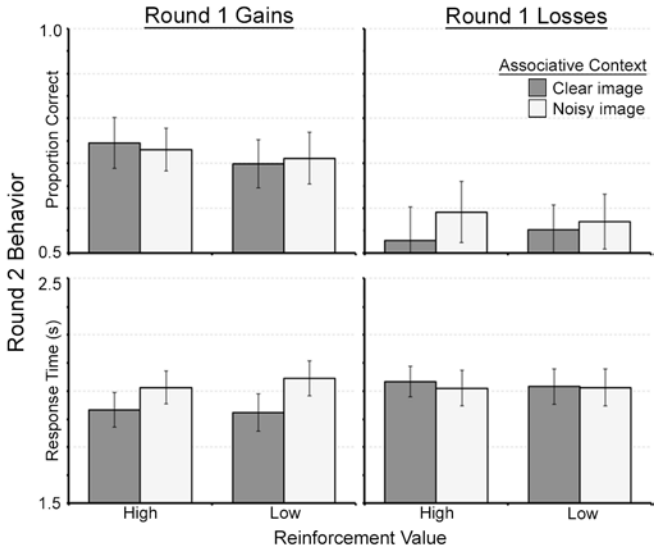

Figure 3. Round 2 decision behavior.

Average accuracy (top graphs) and response time (bottom graphs) during Round 2 are plotted by factors from Round 1 of the deterministic decision-making task. These factors include Round 1 outcome (gain or loss, in columns), reinforcement magnitude (low or high, x-axis), and associative context (clear or noisy images, grey and white bars, respectively). Error bars reflect standard error of the mean.

2.2.1.2 Surprise memory test

Following the scan session, subjects participated in a surprise memory test that interrogated how well they could recognize and recall event-specific details about individual items from the deterministic decision-making task. Subsequent memory behavior was categorized to describe whether subjects were able to correctly recognize and recall details about an item ("Episodic"), just recognize an item ("Familiarity"), or failed to remember an item entirely ("No Memory"). On average, 50.76% (SE = 5.18, N = 731) of the items were both correctly recognized and correctly associated with episodic details from Round 1. This indicates that subjects generally retained episodic memories for about half of the experienced items. For 28.68% (SE = 3.53, N = 413) of the items, subjects failed to correctly identify episodic details, but were able to correctly recognize the item, indicating familiarity with these items. The remaining 20.56% (SE = 3.51, N

= 296) of items were unrecognized, indicating that subjects had no memory for these items. Additionally, subjects false alarmed on 14.10% (SE = 3.50, N = 189) of the new items, indicating a minimal response bias to respond "old" to the recognition memory probe. Table 1 enumerates the factorial breakdown of subsequent memory behavior.

Table 1. Behavioral memory performance

	Round 1 Gain				*Round 1 Loss*			
Context	Clear		Noisy		Clear		Noisy	
Magnitude	High	Low	High	Low	High	Low	High	Low
No Memory	0.14	0.22	0.20	0.24	0.17	0.21	0.24	0.24
Familiarity	0.24	0.25	0.20	0.22	0.31	0.35	0.38	0.36
Episodic	0.62	0.53	0.60	0.54	0.52	0.44	0.38	0.40

Mean proportion of responses to the subsequent recognition and episodic memory probes in a post-scan surprise memory test. "No Memory" reflects items that were unrecognized by subjects. "Familiarity" reflects items that were correctly recognized by subjects, but with errors in episodic memory. "Episodic" reflects items that subjects correctly recognized and correctly recalled details about. These categories of memory accuracy are separated by three factors from Round 1 of the deterministic decision-making task including outcome (gain or loss), associative context (clear or noisy image), and reinforcement magnitude (high or low monetary reward).

2.2.2 Imaging Results

2.2.2.1 Functional region identification

This study sought to test predictions about how the MTL and the striatum each contribute to deterministic decision-making. In order to identify regions of interest in the MTL and striatum, regions were defined using data from prior work with a concurrent discrimination learning task (Tremel et al., 2016; Tremel et al., 2018). Regions were identified based on this dataset to mitigate circularity in the analysis. The search space for these regions of interest was constrained to an anatomical area encompassing the MTL and the striatum, since the *a priori* hypotheses focused on those key systems alone. Using the first two rounds of the concurrent discrimination

task, seven regions were identified between the MTL (n = 4) and striatum (n = 3), including the hippocampus, caudate nucleus, putamen, and parahippocampal gyrus (Figure 4). Bilateral homologues (e.g., head of the caudate nucleus) were collapsed and considered single regions of interest if their coordinates in each dimension (*x*, *y*, *z*) were within 3 millimeters (using the absolute value of the *x* coordinate since sign indicates hemisphere). All of these regions were functionally active during the concurrent discrimination task and during the present deterministic decision-making task (i.e., exhibited a main effect of time). Single-trial time series were extracted from each region.

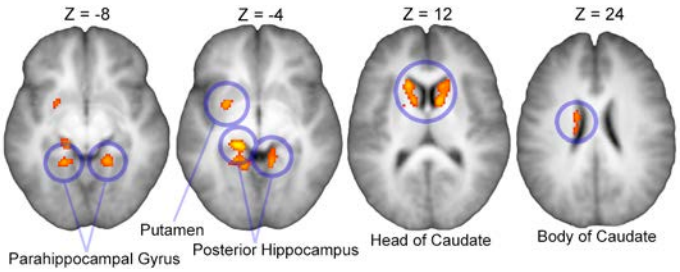

Figure 4. Functional regions of interest in the MTL and striatum.

Magnitude of activation (as z-transformed F-statistic) in the MTL and striatum is plotted on axial slices of the group average, Talairach-transformed brain. These statistics were derived from the main effect of time using data from prior work with a concurrent discrimination learning task (Tremel et al., 2016, 2018). Contiguous voxels in these statistical clusters were considered regions of interest. All regions of interest also exhibited a main effect of time for the present deterministic decision-making task. Red voxels indicate a minimum z-statistic of 4.0, while yellow voxels indicate a maximum z-statistic of 12. Numbers above the axial slices indicate the corresponding z-plane coordinate (millimeters from center plane) in Talairach space.

Activity in four of these regions (Table 2) correlated with accuracy and response time measures, reported in sub-section 2.2.2.2. These regions included the right hippocampus, bilateral parahippocampal gyrus, left putamen, and bilateral head of the caudate nucleus. However, activity in three regions (Table 3) was unassociated with Round 2 decision accuracy and response time. These three regions were excluded from further analysis.

Table 2. Functional regions of interest in the MTL and striatum.

Region	*Anatomical locus*	*x*	*y*	*z*	*Voxels*
1	Right hippocampus	9	-45	0	129
2	Bilateral parahippocampal gyrus	-17/18	-47/-46	-7/-10	72/87
3	Left putamen	-24	0	-5	49
4	Bilateral head of caudate nucleus	-14/11	7/8	-14/-11	254/227

These are the primary regions of interest of this study. *Anatomical locus,* the approximate anatomical location of the center of mass of each region; *x, y, z,* Talairach atlas space coordinates of the center of mass for each region. Two sets of coordinates are reported for bilateral regions indicating the center of mass for the left and right homologues (collapsed when coordinates are within 3 mm). *Voxels,* size of each region in number of voxels (2 mm isotropic voxels).

Table 3. Other regions of interest unassociated with decision-making measures.

Region	*Anatomical locus*	*x*	*y*	*z*	*Voxels*
5	Left hippocampus	-16	-34	-5	80
6	Left body of caudate nucleus	-15	-12	24	34
7	Left parahippocampal gyrus	-10	-51	-2	61

These regions of interest were unassociated with decision-making measures (reported in sub-section 2.2.2.2) and excluded from further analysis. *Anatomical locus*, the approximate anatomical location of the center of mass of each region; *x, y, z,* Talairach atlas space coordinates of the center of mass for each region. Two sets of coordinates are reported for bilateral regions indicating the center of mass for the left and right homologues (collapsed when coordinates are within 3 mm). *Voxels,* size of each region in number of voxels (2 mm isotropic voxels).

2.2.2.2 Dissociating the differential roles of the MTL and striatum in decision-making

The first goal of this study was to identify how the MTL and striatum differentially contribute to deterministic decision-making. It was predicted that both systems would be engaged during decision-making and that activity from both systems would predict behavioral metrics such as accuracy and RT. To establish roles for each of the identified regions of interest, a mixed-effects regression analysis of the Round 2 imaging data was performed to connect to specific trial-level decision behavior. One regression analysis leveraged BOLD activity to predict decision accuracy as a function of Round 1 Outcome (i.e., whether the item was associated with a gain or loss in Round 1) using logistic regression. A second analysis leveraged activity to predict decision speed

(RT) as a function of Round 1 Outcome using linear regression. While these models have four possible effects (intercept, activity, Round 1 Outcome, and activity × Round 1 Outcome), this study will focus on effects related to regional BOLD activity since the goal is to link activity to decision behavior. Full results of each regression analysis are listed in Appendices A and B.

Activity in two regions predicted Round 2 decision accuracy during the deterministic decision-making task. First, activity in the right hippocampus had a negative relationship to trial-level decision accuracy, $\beta = -0.29$, $SE_\beta = 0.13$, $z = -2.23$, $p < 0.05$. A unit increase in hippocampal activity above the mean was associated with a 25% decrease in the odds of making a correct decision in Round 2 versus making an error (Figure 5). In other words, positive hippocampal activity was associated with an increased propensity to make an error in Round 2, while negative activity was associated with an increased propensity to make a correct decision. Regardless, the magnitude of activity (i.e., strong negative-going activation) corresponded to increases in the odds of making a correct response. To quantify the quality of fit for these logistic models, concordance and discordance metrics were computed (Table 4).

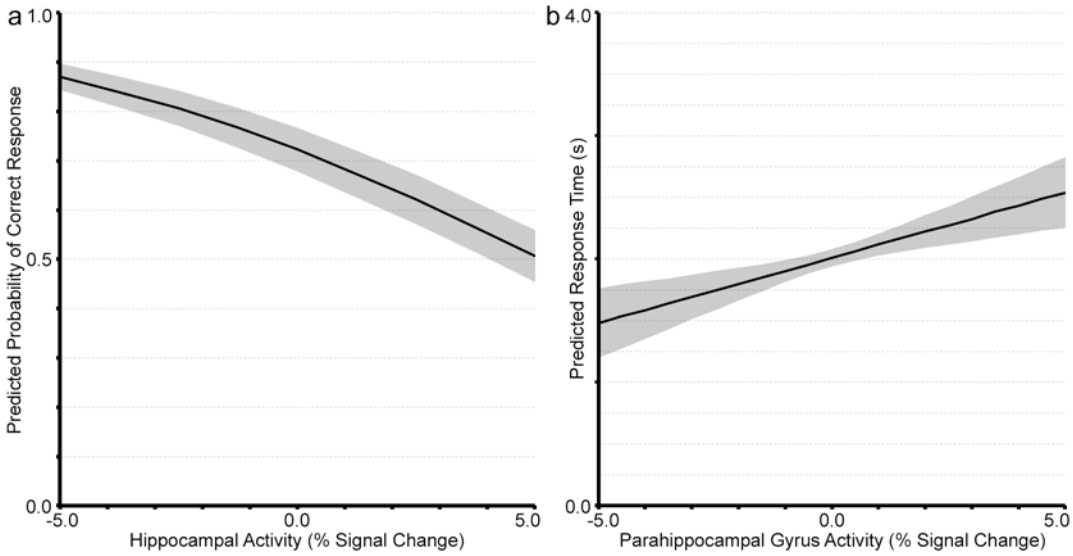

Figure 5. MTL activity predicts decision accuracy and response time

Mixed-effects regression analyses revealed that activity of regions in the MTL predicted measures of decision-making in Round 2 of the deterministic decision-making task. (a) Activity in the right hippocampus predicted decision accuracy in a mixed-effects logistic regression. The predicted probability of making a correct decision in Round 2 (derived from the regression) is plotted against hippocampal activity (as % signal change from baseline). (b) Activity in the bilateral parahippocampal gyrus predicted decision-making speed in a mixed-effects linear regression. The predicted response time (in seconds) of a decision in Round 2 is plotted against activity in the parahippocampal gyrus region (as % signal change from baseline). The shaded area reflects standard error of the mean.

Table 4. Quality of regression model fits

	Accuracy models		*RT models*	
Region	Concordance	Discordance	Conditional r^2	Marginal r^2
R Hippocampus	0.78	0.22	0.23	0.004
Bilat. parahippocampal G.	0.79	0.21	0.23	0.01
L Putamen	0.78	0.22	0.22	0.003
Bilat. head of caudate	0.78	0.22	0.24	0.02

For the mixed-effects logistic regression models (*Accuracy models*), concordance and discordance measures were computed to quantify the quality of model fits. Concordance reflects when the predicted probability of the expected outcome (based on the model) is higher than that for the outcome not occurring. Discordance reflects the opposite. For the mixed-effects linear regression models (*RT models*), conditional and marginal r^2 values were computed to quantify model fit quality. Conditional r^2 reflects the proportion of variance attributable to the fixed and random effects of the model, while marginal r^2 reflects the proportion of variance due to the fixed effects alone. *R*, right; *Bilat*, bilateral; *G*, gyrus; *L*, left.

In a second regression analysis, a different MTL region in the bilateral parahippocampal gyrus was associated with decision speed (RT) during Round 2 of the deterministic decision-making task. Activity in this region predicted decision speed, $\beta = 0.14$, $SE_\beta = 0.06$, $t = 2.47$, $p < 0.05$, such that an increase in RT such that a one-unit increase in activity added 0.14 s to the overall response time (Figure 5). Fit qualities of these linear regression models are enumerated in Table 4.

Parallel to these findings, two regions in the striatum were associated with decision accuracy and with decision speed. First, activity in the left putamen predicted decision accuracy in Round 2, but did so in a manner that depended on Round 1 outcome (gain or loss during initial experience), interaction: $\beta = -0.49$, $SE_\beta = 0.21$, $z = -2.37$, $p < 0.05$. This interaction with Round 1 outcome indicated that activity in the putamen coded for the learned valence of the item (i.e., whether it was a positive gain or negative loss). For gain items, higher activity was associated with an increase in the odds of making a correct decision, while for loss items, lower activity (i.e., more negative) was associated with an increase in the odds of making a correct decision (Figure 6).

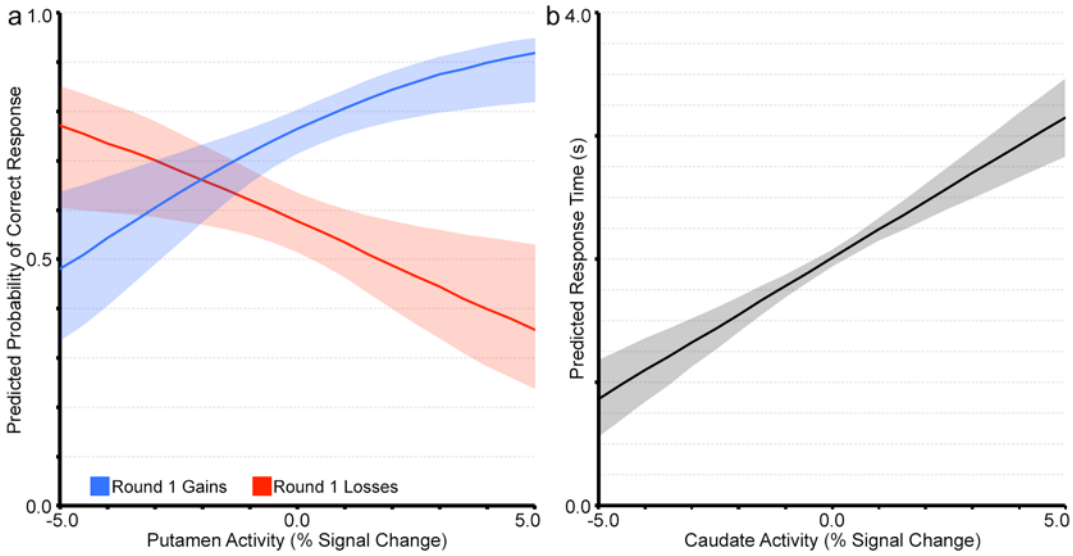

Figure 6. Striatal activity predicts decision accuracy and response time

Mixed-effects regression analyses revealed that activity of regions in the striatum predicted measures of decision-making in Round 2 of the deterministic decision-making task. (a) Activity in the left putamen predicted decision accuracy in a manner dependent on Round 1 outcome (gain, loss) in a logistic regression. The predicted probability of making a correct decision in Round 2 (based on the model) is plotted against putamen activity (as % signal change from baseline). Blue line indicates effect for trials associated with monetary gain in Round 1, while the red line indicates the effect for trials associated with monetary loss in Round 1. (b) Activity in the bilateral head of the caudate predicted decision-making speed in a linear regression. The predicted response time (in seconds) of a decision in Round 2 is plotted against activity in the caudate nucleus. The shaded areas reflect standard error of the mean.

In the second regression analysis, activity in the bilateral head of the caudate nucleus (region 3) predicted trial-level decision RT, $\beta = 0.21$, $SE_\beta = 0.07$, $t = 3.02$, $p < 0.01$ (Figure 6). Activity was positively related to RT such that for every unit-increase in activity, RT increased by an average of 0.21 s.

Altogether, activity during Round 2 in the MTL and striatum correlated with measures of Round 2 decision behavior. Specifically, activity in the right hippocampus was anti-correlated with decision accuracy, while activity in the left putamen predicted decision accuracy in a manner dependent on initial decision outcome (i.e., gain or loss in Round 1). Activity in the bilateral parahippocampal gyrus and the bilateral head of the caudate nucleus seemed to

contribute to decision speed, wherein engagement of these regions was associated with increases in the length of a decision episode. Thus, for the MTL system and the striatal system, one region in each predicted decision accuracy (right hippocampus, left putamen) and another region predicted decision speed (bilateral parahippocampal gyrus, bilateral caudate head).

2.2.2.3 Characterizing the role of the MTL and striatum in memory encoding

While tying MTL and striatal activity to Round 2 decision-making behavior establishes that both systems functionally contribute to decisions, it is unclear whether these systems contribute in fundamentally different ways. The second goal of this chapter was to further characterize the functional roles of these systems by examining how regional activity during the initial decision experience (i.e., Round 1) related to post-task measures of subsequent memory. It was expected that the MTL would contribute instance-specific mnemonic information. Therefore, Round 1 activity in MTL regions should be predictive of later episodic memory performance wherein subjects were asked to recall experience-related information about particular items. This relationship was expected to be absent for striatal regions.

To test this hypothesis, regional activity during Round 1 was assessed based on subsequent recognition and episodic memory performance. Items were sorted according to behavioral responses to the subsequent memory probes and entered as random factors into ANOVAs. Two ANOVAs were computed per region for the recognition and episodic memory probes, respectively. Given that the goal of this analysis was to examine how decision-related MTL and striatal regions behaved during initial (Round 1) memory encoding, regions that failed to exhibit connections to subsequent decision-making behavior were dropped from further analysis. These included regions in the left hippocampus, left body of the caudate nucleus, and left parahippocampal gyrus (Table 3).

The first analysis assessed item-level encoding activity based on the behavioral performance of subjects to the surprise recognition memory test that followed the deterministic decision-making task. This ANOVA tested Round 1 activity by four factors, including one derived from subsequent memory accuracy (hit, miss), and three from the initial encoding experience of the deterministic decision-making task. These factors included Round 1 outcome (gain, loss), reinforcement magnitude (high, low), and associative context (clear, noisy). The full ANOVA results are enumerated in Appendix C. One MTL region and one striatal region exhibited statistically significant effects associated with recognition memory accuracy (i.e., hits or misses).

Specifically, in the right hippocampus, there was a significant interaction between memory accuracy (hit, miss) and reinforcement magnitude (high, low), $F[1, 1370] = 8.93$, $p < 0.05$ (Figure 7). Activity for hit items was equal regardless of reinforcement magnitude, $t[1141.8] = 0.11$, $p = 0.91$, but for miss items, activity was greater for low magnitude misses than for high magnitude misses, $t[231.8] = -2.70$, $p < 0.01$. Within each level of magnitude, activity during encoding for high magnitude hits was greater than that for high magnitude misses, $t[135.84] = -3.20$, $p < 0.01$, but there was no such difference for low magnitude hits and misses, $t[114.36] = 0.03$, $p = 0.97$. It is worth noting that activity in the hippocampus was positive here during encoding, whereas it was negative during Round 2 decision-making. This is consistent with findings illustrating positive activity in the MTL during encoding and negative activity during memory retrieval (Davachi & Wagner, 2002; Suzuki et al., 2011; Yu et al., 2012).

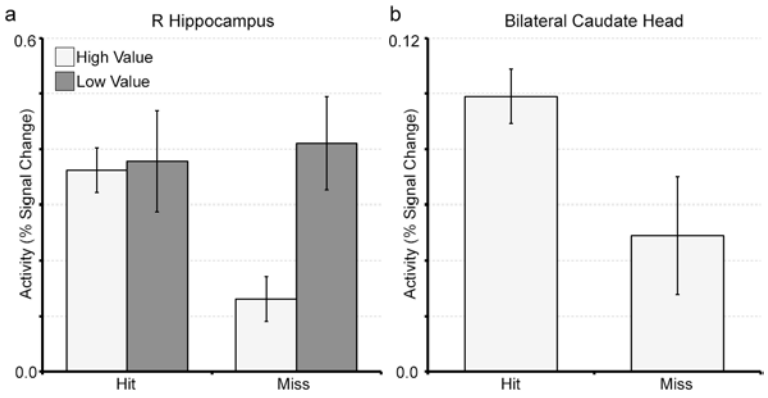

Figure 7. MTL and striatal activity predict recognition memory performance

(a) Mean activity in the right hippocampus is plotted for items that were later correctly recognized (hits) or unrecognized (misses) in a post-scan surprise memory test. Items associated with high value monetary gain or loss in Round 1 (high value) is plotted separately from those items associated with low value monetary gain or loss to illustrate the interaction. (b) Mean activity in the bilateral caudate head is plotted for hit and miss items based on the post-scan recognition memory test.

A second analysis assessed item-level encoding activity based on the behavioral performance of subjects to the surprise episodic memory test that followed the deterministic decision-making task. This ANOVA tested whether Round 1 activity was sensitive to experimental factors and episodic memory behavior. One factor, memory response, was derived from subsequent memory performance. Items were categorized based on the behavioral responses of subjects describing whether they thought they received positive or negative feedback during Round 1 of the decision task (i.e., "gain" responses versus "loss" responses). Three other factors were included, derived from the initial encoding experience of the decision task, including Round 1 Outcome (gain, loss), reinforcement magnitude (high, low), and associative context (clear, noisy). The full ANOVA results are listed in Appendix D. Two regions in the MTL exhibited statistically significant effects associated with behavioral responses to the episodic memory probe, but neither striatal region exhibited an effect.

In the right hippocampus, there was a significant interaction predicting episodic memory accuracy, $F[1, 1292] = 13.84$, $p < 0.01$ (Figure 8). For items associated with monetary gain in Round 1, activity was higher when items were successfully remembered than for those incorrectly remembered, $t[466.58] = 2.67$, $p < 0.01$. This pattern was also present for items associated with losses in Round 1, $t[506.42] = 3.05$, $p < 0.01$, suggesting that activity in the hippocampus during encoding reflected subsequent episodic memory accuracy.

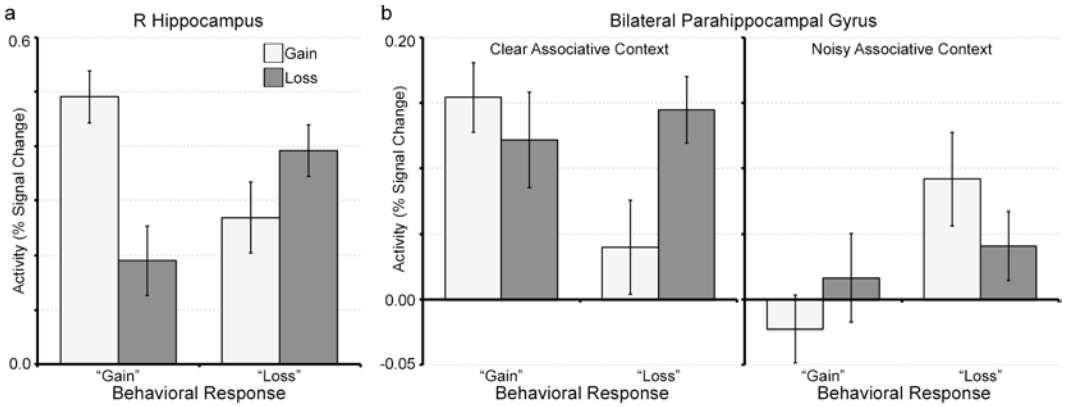

Figure 8. MTL activity predicts episodic memory ability.

(a) Mean activity in the hippocampus is plotted based on the behavioral response of subjects to the post-scan surprise episodic memory test. Items that subjects believed were associated with monetary gains or losses are labeled on the x-axis. Whether or not the item was actually associated with a monetary gain or loss are plotted as separate light or dark grey bars, respectively. Accurate episodic memory judgments correspond to the "Gain" + Gain and "Loss" + Loss categories. (b) Mean activity in the parahippocampal gyrus is plotted based on behavioral response of subjects to the post-scan surprise episodic memory test as in panel (a). Data are plotted separately for items associated with a clear context (i.e., clear image in Round 1) versus a noisy context (i.e., noisy image in Round 1) to illustrate the interaction.

In the bilateral parahippocampal gyrus, there were two statistically significant interactions. There was an interaction of memory response ("gain," "loss" responses) and associative context, $F[1, 1292] = 8.38$, $p < 0.05$. This interaction was also modulated by a third factor of Round 1 outcome, $F[1,1292] = 8.97$, $p < 0.05$ (Figure 8). For the clear context condition, parahippocampal activity during encoding was associated with episodic memory

accuracy, wherein activity was greater for items that were successfully remembered for both Round 1 gains, $t[259.29] = 2.70$, $p < 0.01$, and Round 1 losses, $t[229.82] = -2.02$, $p = 0.04$. This pattern held based on behavioral response to the memory probe, including "gain" responses, $t[244.97] = 2.11$, $p = 0.04$, and "loss" responses, $t[243.92] = -2.39$, $p = 0.02$. This indicates that the parahippocampal gyrus acts similarly to the hippocampus with respect to episodic encoding for clear context items. However, for the noisy context condition, encoding-related activity was related only to subjects' behavioral responses to the memory probe, wherein "loss" responses were associated with higher activity than "gain" responses for items associated with monetary gains in Round 1, $t[222.47] = -2.45$, $p = 0.01$, and for items associated with monetary losses in Round 1, $t[250.03] = -2.04$, $p = 0.04$. Within each of these behavioral response categories, activity did not differ by Round 1 outcome (i.e., whether the item was actually associated with a gain or loss in Round 1) for "gain" responses, $t[259.14] = -0.49$, $p = 0.63$, nor for "loss" responses, $t[214.72] = 0.82$, $p = 0.41$. This pattern indicates that activity in this region encoded subsequent behavior to the episodic memory probe for noisy context items, but this encoding does not reflect the veracity of what was actually experienced in Round 1 (i.e., Round 1 outcomes).

To summarize, activity during the initial decision experience in the right hippocampus and bilateral caudate both predicted recognition memory performance. Additionally, activity in the right hippocampus and bilateral parahippocampal gyrus both predicted episodic memory performance. This suggests that both MTL and striatal systems may encode information that supports the ability to identify that a particular stimulus had been encountered before. However, the MTL alone seems to support the recall of instance-specific information about previous decision episodes.

2.2.2.4 Functional connectivity between MTL, striatum, and ventral stream regions

The final goal of these analyses was to examine the functional relationships between the MTL and the striatum and to further characterize their relationship with stimulus-selective visual regions involved in the processing of stimuli from a Round 1 decision experience. Given that both systems were found to contribute to decision behavior but associated with distinct memory profiles, it was expected that the MTL and striatum would operate in parallel. In other words, each system supplies key information to the decision process with minimal interaction between them. It was also expected that each system would differentially respond to the 2 x 2 factorial manipulation of associative context and reinforcement magnitude in Round 1, with the MTL exhibiting more sensitivity to the context manipulation and the striatum exhibiting more sensitivity to the manipulation of reinforcement magnitude.

To examine these putative relationships, two functional connectivity analyses were performed, testing first how the relationship between MTL, striatal, and stimulus-selective visual regions modulated by associative context, and second how connectivity modulated by reinforcement magnitude. Functional connectivity was computed using each of the four regions as seeds in separate analyses. This was operationalized as a psychophysiological interaction (PPI) from a seed region to each of the target regions. A PPI tests whether psychological factors (e.g., associative context or reinforcement magnitude) influence the coupling of underlying physiological signals between two regions. A positive PPI effect indicates that functional connectivity between two regions is modulated by a particular factor from the task.

Stimulus-selective visual regions defined from the localized task were included in this analysis to test the extent to which MTL and striatal regions may draw upon representations of instance-specific information from the initial decision experience. For example, the MTL may

encode visual information about specific images that appeared in Round 1. If this is the case, MTL regions should exhibit increased functional connectivity to scene-selective parahippocampal gyrus regions for the clear context condition versus the noisy context condition.

For each seed region, I first assessed the extent to which associative context manipulation (clear versus noisy images) altered functional connectivity. Second, I assessed the impact of reinforcement magnitude (low or high) on connectivity. These factors were computed as a psychological (i.e., task) regressor within the PPI, convolved with a hemodynamic response function to account for offset due to BOLD signal lag. For the stimulus-selective visual region targets, three regions were localized, including two high-order visual processing regions selective for faces or scenes and one word-selective region based prior studies (Cohen et al., 2000; Cohen et al., 2002; McCandliss et al., 2003) (Table 5).

Table 5. Stimulus-selective regions of interest in the ventral stream

Region	*Anatomical locus*	*x*	*y*	*z*	*Voxels*	*Selectivity*
1	Bilateral parahippocampal gyrus	26/-27	-52/-62	-8/-9	482/-368	Scenes
2	Right fusiform gyrus	38	-54	-15	209	Faces
3	Left fusiform gyrus (VWFA)*	-45	-57	-11	155	Words

These additional regions of interest were identified using a localizer task. The asterisk for the VWFA region indicates that this region was defined using coordinates from the literature. Notably, this region exhibited selectivity for word versus face stimuli in the present localizer. *Anatomical locus*, approximate anatomical location of the center of mass of the region; *x*, *y*, *z*, Talairach atlas space coordinates of the center of mass of the region (bilateral region has two sets of coordinates corresponding to the left and right homologues); *Voxels*, size of the region in voxels (2 mm isotropic); *Selectivity*, statistical effect exhibited by the region (e.g., "Scenes" indicates that this region responds preferentially to images of scenery).

The first PPI analysis examined changes in functional connectivity as a function of the associative context manipulation (Figure 9). Regions in the MTL were predicted to be especially

sensitive to this factor, since subjects could use the clear background images to encode a unique and vivid episodic decision experience versus a noisy background image. Functional connectivity was expected to be greater for MTL regions to other regions, especially to the scene-selective region in the parahippocampal gyrus. Indeed, functional connectivity was greater for clear contexts versus noisy contexts between the right hippocampus and bilateral parahippocampal gyrus ($\beta = 0.02$, $r = 0.26$, $p < 0.001$). This was also true from the right hippocampus to the scene-selective parahippocampal region ($\beta = 0.02$, $r = 0.11$ $p < 0.001$) and from the bilateral parahippocampal gyrus to the scene-selective region ($\beta = 0.04$, $r = 0.23$, $p < 0.001$). Connectivity from the right hippocampus to the approximate VWFA also increased for clear images versus noisy images ($\beta = 0.01$, $r = 0.02$, $p < 0.05$), suggesting that hippocampal processing may rely on visual representations of word stimuli.

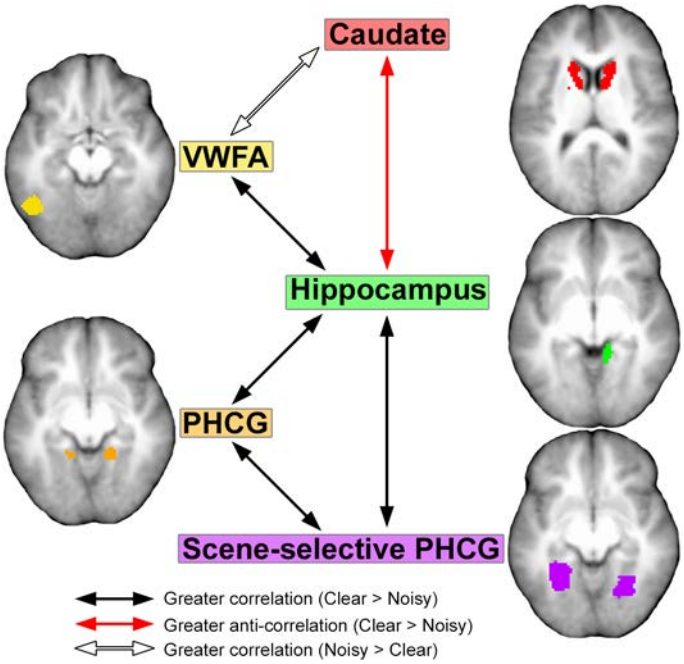

Figure 9. Functional connectivity of the MTL and striatum modulates with context.

Statistically significant functional connectivity effects dependent on the associative context manipulation (defined by a psychophysiological interaction) are plotted as arrows between regions of interest. Black arrows represent greater functional connectivity between two regions for clear contexts (i.e., clear images) than for noisy contexts. Red arrows represent greater negative functional connectivity (i.e., anti-correlation of activity) between two regions for clear versus noisy contexts. White arrows represent greater functional connectivity for noisy versus clear contexts. *VWFA*, visual word form area; *PHCG*, parahippocampal gyrus.

Additionally, functional connectivity from the bilateral head of the caudate nucleus to the right hippocampus also modulated with associative context. The right hippocampus and bilateral caudate exhibited negative functional connectivity ($\beta = -0.03$, $r = 0.08$, $p < 0.01$), wherein there was greater anti-correlation between regions for the clear context versus the noisy context. This suggests that engagement of the hippocampus may be emphasized during clear-image trials relative to that of the caudate. Further supporting this, the bilateral caudate head also exhibited a PPI with the approximate VWFA ($\beta = -0.06$, $r = 0.06$, $p < 0.01$), such that connectivity increased for noisy contexts relative to clear contexts.

In a second PPI analysis, modulations in functional connectivity as a function of reinforcement magnitude were examined (Figure 10). The striatum was expected to be especially sensitive to this factor, since this system conventionally responds to value manipulations (Delgado et al., 2000; Elliott et al., 2000; Liu et al., 2007; Valentin et al., 2007). Functional connectivity from the bilateral head of the caudate nucleus to the left putamen increased for low magnitude trials relative to high magnitude trials ($\beta = 0.15$, $r = 0.12$, $p < 0.05$). The caudate again exhibited a negative relationship to the right hippocampus ($\beta = 0.96$, $r = 0.04$, $p < 0.01$), wherein there was greater anti-correlation between regions for low magnitude trials versus high magnitude trials. This suggests that engagement of the striatum may be emphasized on high magnitude trials relative to the hippocampus (which exhibited preference for low magnitude trials). For the right hippocampus, functional connectivity increased for low versus high magnitude trials to both the bilateral parahippocampal gyrus region ($\beta = 0.01$, $r = 0.26$, $p < 0.05$) and the scene-selective region ($\beta = 0.01$, $r = 0.09$, $p < 0.01$).

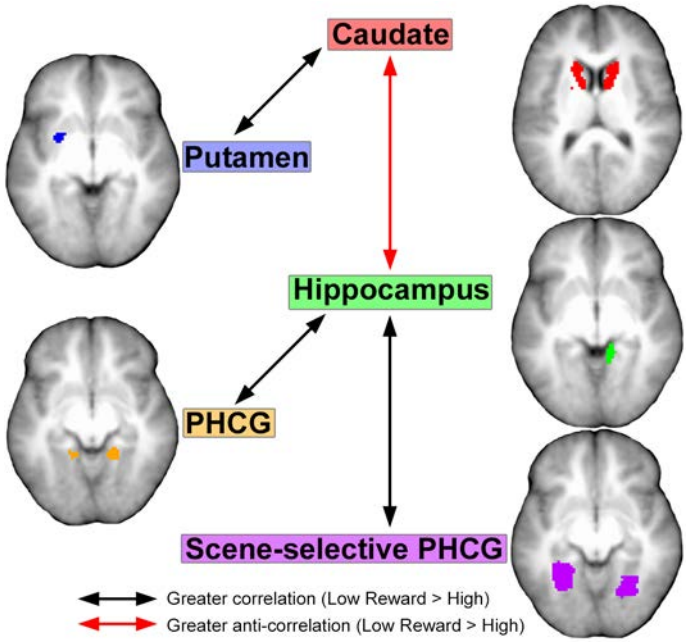

Figure 10. Functional connectivity of the MTL and striatum modulates with magnitude.

Statistically significant functional connectivity effects dependent on the reinforcement magnitude manipulation are plotted as arrows between regions. Black arrows represent greater functional connectivity between two regions for low value reinforcement trials versus high value reinforcement trials. Red arrows represent greater negative functional connectivity (i.e., anti-correlation of activity) for low value versus high value trials. *PHCG*, parahippocampal gyrus.

2.2.2.5 Summary of findings

The general pattern of results indicates that both the MTL and striatum contribute to deterministic decision-making, but that they do so in fundamentally different ways (Table 6). Activity in the right hippocampus predicted decision accuracy during Round 2 and subsequent memory performance. Additionally, functionally connectivity from the hippocampus was sensitive to both associative context and reinforcement magnitude. The hippocampus was functionally connected to both the VWFA and scene-selective regions, suggesting that it may draw upon or encode visual representations from the initial decision episode.

Table 6. Summary of results

Region	*Decision Effects*	*Encoding Effects*	*Connectivity Effects*
Hippocampus	• Accuracy	• Recognition memory accuracy (high value items) • Episodic memory accuracy	• (+) PHCG (clear context > noisy) • (-) Caudate (clear context > noisy) • (+) Scene (clear context > noisy) • (+) VWFA (clear context > noisy) • (+) PHCG (low reward > high) • (-) Caudate (low reward > high) • (+) Scene (low reward > high)
PHCG	• Speed	• Episodic memory accuracy (clear context items)	• (+) HC (clear context > noisy) • (+) Scene (clear context > noisy) • (+) HC (low reward > high)
Putamen	• Accuracy (by gain, loss)	• N/A	• (+) Caudate (low reward > high)
Caudate head	• Speed	• Recognition memory accuracy	• (-) HC (clear context > noisy) • (+) VWFA (noisy context > clear) • (-) HC (low reward > high) • (+) Putamen (low reward > high)

Major effects for each region are enumerated in three categories. *Decision Effects*, association between regional activity in Round 2 and decision measures (accuracy and response time) during Round 2 of the deterministic decision-making task. *Encoding Effects*, association between regional activity during Round 1 of the decision-making task and memory measures from the post-scan surprise memory test. *Connectivity Effects*, functional connectivity effects, including direction and psychological factor, between region at the left of the table to target region listed in this column. *(+)* indicates a direct correlation in activity of the two regions, while *(-)* indicates an anti-correlation in activity of the two regions. Psychological factor underlying the effect (associative context or reinforcement magnitude) and the direction of the effect are indicated in parentheses. *PHCG*, parahippocampal gyrus; *VWFA*, visual word form area; *HC*, hippocampus.

Another MTL region, the bilateral parahippocampal gyrus, was associated with decision speed and episodic memory performance. Functional connectivity from this region to the right hippocampus was modulated by both associative context and reinforcement magnitude. Additionally, the parahippocampal gyrus exhibited associative context-dependent connectivity to a neighboring scene-selective visual region, suggesting that this region may encode information about the visually salient images in Round 1.

In the striatum, activity in the left putamen predicted decision accuracy in Round 2 in a manner dependent on Round 1 outcome (i.e., gain or loss), suggesting it may encode decision-relevant value information. Consistent with this, functional connectivity from the left putamen to the bilateral caudate head modulated with reinforcement magnitude. Activity in the bilateral head

of the caudate nucleus predicted decision speed and recognition memory performance. It was functionally connected to the VWFA in a manner dependent on associative context, suggesting that it may preferentially encode stimulus-associated information in situations with indistinct contexts (e.g., stimulus-outcome relationships). Additionally, activity in the caudate was anti-correlated with that in the right hippocampus dependent on both associative context and reinforcement magnitude, suggesting that these two systems may operate in parallel. For instance, when the striatal system was most active, the hippocampal system seemed to be less active.

2.3 DISCUSSION

In this study, I investigated the differential contributions of the MTL and striatum in a single-exposure deterministic decision-making task. To manipulate the relative engagement of these systems, this task featured a 2 x 2 factorial manipulation of associative context (via images) and reinforcement magnitude (via money) related to the individual decision experiences. Activity of regions in the MTL and striatum predicted decision behavior and subsequent memory ability. Specifically, the hippocampus was linked to episodic memory and decision accuracy, while the parahippocampal gyrus was linked to contextual aspects of episodic memory and decreases in decision speed. In the striatum, the caudate was associated with recognition memory performance and decision speed, while the putamen was associated with decision accuracy, but not measures of subsequent memory. Functional connectivity in the MTL increased for high context trials, while that in the striatum increased for low context trials. Connectivity between the MTL and the caudate nucleus modulated by context and by reinforcement magnitude.

Collectively, these findings characterize the neural basis underlying the mnemonic representations that drive single-exposure deterministic decision-making.

2.3.1 MTL regions support the recall of instance-specific information to guide decisions

A region in the right posterior hippocampus was found to support the encoding of episodic information that can be used to drive subsequent decision-making after a single exposure. Activity in this region during the initial encoding experience (Round 1) predicted later recognition memory and episodic memory retrieval success. This suggests that the hippocampus encodes episodic information about individual instances, for example what word was chosen, what image was associated with that experience, or what outcome was received (Corkin, 2002; Davachi & Wagner, 2002; Eichenbaum, 2001; Squire & Zola, 1996; Squire & Zola-Morgan, 1991). This information is critical for producing subsequent decision behavior, wherein an individual may reconstruct the initial episode from these details to reason about which stimulus is best to choose. Indeed, hippocampal activity in Round 2, presumably reflecting the retrieval of these episodic details (Davachi & Wagner, 2002; Squire & Zola-Morgan, 1991), coded for decision accuracy, suggesting that individuals rely on instance-specific, episodic information to drive decisions after a single exposure. This is consistent with other findings implicating the hippocampus as a key substrate in supporting subsequent decision behavior (Bornstein & Daw, 2013).

While this finding may be unsurprising in that the hippocampus is thought to support rapid learning (Euston et al., 2012; Kumaran et al., 2009), it illustrates that the MTL may serve a critical role early in the learning trajectory. If the MTL is critical for supporting the first few decision-making opportunities in a repetitive decision-making task, then damage to this region

should impair learning. Neuropsychological findings support this assertion, wherein MTL damage typically impairs or drastically slows deterministic learning (Bayley et al., 2005; Buffalo et al., 1998; Buffalo et al., 1999; Chudasama et al., 2008; Corkin, 2002; Gaffan & Murray, 1992; Hood et al., 1999; Malamut et al., 1984; Phillips et al., 1988; Squire et al., 1988; Squire and Zola, 1996; Suzuki et al., 1993; Zola-Morgan et al., 1989; Zola-Morgan et al., 1994).

Alongside the hippocampus, other MTL regions, such as the parahippocampal gyrus, may supplement this primary role by storing and providing additional associative details about prior experiences. In the present study, activity in the parahippocampal gyrus predicted episodic memory retrieval success specifically for items that were paired with a clear (i.e., not noise-degraded) natural landscape image. This suggests that the parahippocampal gyrus may be important for encoding contextual information about an experience, such as visual imagery. This may be why functional connectivity between the parahippocampal gyrus, scene-selective regions, and the hippocampus modulated with this associative context manipulation as well, such that connectivity among the three regions was greater for trials associated with clear images.

The parahippocampal gyrus was also associated with subsequent decision-making behavior in Round 2 of the task, wherein increases in activity predicted increases in decision response times (i.e., slower decisions). These increases in decision times might result from the recollection of additional details, such as reconstructing associated visual imagery, related to parahippocampal engagement. This context-dependent recruitment of the parahippocampal gyrus has been linked to the processing of associative information that is bound to information about items (e.g., words in the present task) in the hippocampus (Howard et al., 2011). Thus, together with the hippocampus, these regions constitute a memory system that can support rapid learning

of detail-driven information, which can improve subsequent decision-making in a short time frame (Euston et al., 2012; Kumaran et al., 2009).

However, the extent of the engagement of these MTL regions may depend on contextual factors surrounding a decision experience. While the parahippocampal gyrus specifically supported context-dependent memory, both hippocampal and parahippocampal regions were preferentially more coupled during rich associative contexts (i.e., clear images) to each other, to scene-selective regions in the ventral stream, and to a putative word-representation region in the fusiform gyrus. This is consistent with findings that adding imagery to decision experiences can modulate the engagement of the MTL as a whole (Bornstein et al., 2012; Bornstein et al., 2013; Doll et al., 2015b; Hannula et al., 2013; Hayes et al., 2010; Howard et al., 2011; Park et al., 2014). Indeed, factors that affect the visual context of an experience have been linked to reactivations of visual representations during MTL-based memory recall (Mack & Preston, 2016). As such, the MTL seems to rely upon the reactivation of these memory traces of the initial experience as a means to inform future decisions.

2.3.2 Sub-regions of the striatum are engaged in single-exposure decision-making

Two sub-regions of the striatum, the putamen and head of the caudate nucleus, were also found to support different aspects of single-exposure deterministic decision-making. The putamen coded for decision accuracy in Round 2 after a single experience with the items, but did so in a manner dependent on item valence. Positive-going activity in this region reflected improved Round 2 decision accuracy for positively valued items (i.e., items associated with monetary gains in Round 1) while negative-going activity reflected improved accuracy for negatively valued items (i.e., items associated with monetary losses in Round 1). This suggests that the putamen

codes for the value associated with a decision outcome, such that experiencing a gain or loss can produce an update to predictions of expected outcomes for future choices (Seger et al., 2010; Tremel et al., 2016). However, this region was unrelated to metrics of subsequent memory, suggesting that while this region may encode information that can be used to inform subsequent decision-making, this region may not support the retrieval of individual instances of experience. Indeed, the mnemonic representations supported by the putamen typically require tuning across several repetitions or experiences and are unlikely to be reliable after just a single exposure (Packard & Knowlton, 2002; Yin et al., 2004). Thus, the putamen may be involved in laying a mnemonic foundation to set up this tuning early on. This foundation, while it can actively contribute to early decision-making, represents a relatively unstable initial prediction about the value of having made a particular choice once. With repetition, this decision value can be refined by compiling a history of experience with a given choice into a single prediction (Seger et al., 2010; Tremel et al., 2016). In some ways, the putamen may serve a parallel role as the hippocampus in this type of learning, wherein both regions predicted accuracy. However, it seems that each region processes fundamentally different types of information, such that the hippocampus can support instance-based retrieval, but the putamen cannot.

The head of the caudate nucleus supported subsequent decision-making behavior in a different way. Greater engagement of this region was correlated with increases in decision times, suggesting that this region may be involved in control-related functions such as cognitive performance evaluation, feedback processing, or the monitoring of goal achievement (Tremel et al., 2016; Tricomi & Fiez, 2008, 2012). In prior work, similar regions in the caudate head were found to correlate with increases in decision thresholds, estimated by drift-diffusion modeling, and with the magnitude of reward prediction errors, estimated by model-free reinforcement

learning agents (Tremel et al., 2016, 2018). This suggests that the caudate may respond primarily to unexpected situations (e.g., an error in outcome prediction) and signal for the need to expend greater effort in making a decision, observed as increases in response times (in the present study) and increases in decision thresholds (in prior work). In this context, increases in decision thresholds can capture increases in decision effort, associated with longer response times (Usher & McClelland, 2001). This increase in effort may reflect several processes, including feedback processing to enable additional encoding of relevant details about an experience (Tricomi & Fiez, 2008) or the updating of underlying value predictions to ensure better future performance (Bornstein & Daw, 2011; Gläscher et al., 2010; Haruno et al., 2004; Kim & Hikosaka, 2013).

This putative association between the caudate and decision effort is underscored by the relationship here between Round 1 caudate activity and subsequent recognition memory accuracy. Increases in effortful processing due to an unexpected outcome can increase the salience of individual items and in turn, increase the ability to subsequently recognize them (Kim & Hikosaka, 2013). In this way, the caudate may serve as some kind of link between multiple memory systems, wherein item-level influences can impact other memory representations. For instance, activity in the caudate modulates with the strength of declarative memory representations (i.e., via the MTL) that are drawn upon during subsequent choices (Tricomi & Fiez, 2012). Moreover, the present study found that the caudate was functionally connected to both the putamen and the hippocampus depending on modulations in two experimental factors. Thus, it seems that the caudate nucleus functions at least partly in the capacity of cognitive control and may coordinate simultaneously with other striatal regions (e.g., the putamen) and the MTL (though perhaps indirectly) to optimize future decisions.

Importantly, both of these striatal regions contributed to subsequent decision behavior after just a single exposure, highlighting the importance of the striatum in short-term learning. These striatal regions constitute a procedural memory system that is traditionally thought to require several repetitions and a longer learning trajectory to optimize the underlying mnemonic information (Daw & Doya, 2006; Dayan & Daw, 2008; O'Doherty et al., 2017). Here, however, the caudate and putamen seem to act similarly as they would for a longer learning trajectory, wherein the caudate supports effortful engagement and the putamen supports mnemonic storage (Tremel et al., 2016, 2018). This striatal system is thought to function by encoding, updating, and storing prediction signals about outcome values that are optimized across multiple repetitions of an experience (Cohen & Ranganath, 2007; Gläscher et al., 2010; Hare et al., 2008). In the present study, subjects begin with no prior knowledge about the outcome relationships, so prediction errors would be large and frequent during the exposure round of the task. If this principle holds true for single-exposure learning, the caudate could be responsible here for signaling prediction errors that could update value representations in the putamen (Camille et al., 2011; FitzGerald et al., 2009; Hare et al., 2008; Padoa-Schioppa & Assad, 2006; Schultz et al., 1997; Schultz, 2013). While this initial foundation of striatal predictions may be too inaccurate or unreliable to individually drive decision-making after just one exposure, it is significant that it seems to be in place after a single experience and seems to support subsequent decision-making to some degree (Miyachi et al., 2002; Seger & Cincotta, 2005).

2.3.3 Multiple memory systems contribute in parallel to deterministic decision-making

Different systems in both the MTL and the striatum contribute to improved decision-making performance after a single exposure. Interestingly, one region in each system predicted decision

accuracy, while another region in each system predicted decision speed. This suggests some degree of functional overlap or redundancy between systems, consistent with assertions that the MTL and striatum may act in parallel to support decision-making (Delgado & Dickerson, 2012; Dickerson et al., 2010; Doll et al., 2015b; Poldrack & Rodriguez, 2004; Shadlen & Shohamy, 2016). Regions in the hippocampus and putamen both predicted subsequent decision accuracy, suggesting that both the MTL and striatum may be capable of individually supporting decision behavior. However, given that both systems were simultaneously active, it seems plausible that information from these systems is weighted and integrated downstream to produce a singular decision output (Delgado & Dickerson, 2012; Shadlen & Shohamy, 2016). The hippocampus may build declarative associations between stimuli, outcomes, and experiences, while at the same time the putamen builds predictions about the expected value of a particular choice option. One critical distinction, however, is that the MTL system seems to support the recollection of individual prior instances, whereas the putamen does not. This suggests that in the case of single-exposure learning, the MTL rather than the striatum may provide more accurate and useful information with respect to decision-making.

This functional overlap may explain why learning can be occasionally preserved in cases of MTL damage. By altering the context of a decision in such a way as to emphasize the use of one system over the other, these deficits due to MTL loss may be reduced via striatal compensation (Bayley et al., 2005; Malamut et al., 1984; Rehbein et al., 2005; Ridley et al., 1989). In the present study, functional connectivity between the caudate and hippocampus was anti-correlated, suggesting that the context of a decision (i.e., situations depending on the experimental factors in this study: reinforcement magnitude or associative context) can influence the relative balance of these systems. For instance, on trials associated with clear imagery (i.e.,

associative context manipulation), activity in the hippocampus and the caudate were strongly anti-correlated, suggesting that the MTL may be more efficacious for these trials due to the extra associative details upon which individuals can anchor their experiences. As such, information from the MTL may be more reliable for these trials than that from the striatum due to its ability to support instance retrieval. Taken together, however, both systems seem to be useful for decision-making, even though their relative contributions to a given choice may be weighted. Indeed, the worst decision performance deficits arise when both MTL and striatal regions are damaged simultaneously (Teng et al., 2000; Turchi et al., 2010).

2.3.4 Conclusions

In this study, I investigated how MTL and striatal substrates individually contribute to deterministic decision-making behavior. This was examined in a single-exposure decision task which separated encoding processes from decision-related processes. The MTL system appeared to be necessary for early success in deterministic decision-making and for retaining subsequent memory for individual decision experiences. A striatal memory system was co-active alongside the MTL system, supporting decision accuracy and speed. Furthermore, the relative engagement of these systems depends on contextual factors of an initial decision experience, wherein additional associative context seems to more greatly engage the MTL system, while reinforcement seems to more greatly engage the striatum. Collectively, this study demonstrates that multiple memory systems underlie deterministic decision-making and that contextual factors surrounding a decision event can influence their relative recruitment and efficacy.

3.0 COMPUTATIONAL MECHANISMS UNDERLYING MTL AND STRIATAL CONTRIBUTIONS TO DETERMINISTIC DECISION-MAKING

When making a decision, multiple mnemonic resources are drawn upon to inform our choices (Delgado & Dickerson, 2012; Doll et al., 2015b; Shadlen & Shohamy, 2016; Tremel et al., 2016). In chapter 2, I illustrated that the MTL and striatum support decision-making behavior in different ways after a single exposure to particular choices. These systems seem to underlie different types of information that can each influence decision-making. The MTL provides episodic details about specific decision events, wherein individual instances can be retrieved to inform subsequent behavior. In contrast, the striatum provides predictions about the value of particular decision outcomes, reflecting an accrued history of experience. While both systems may operate in parallel during single-exposure learning and decision-making, it is less clear whether one system may play a dominant role in this type of behavior. This chapter implements a computational reinforcement learning approach to test the possibility that the MTL is essential to successful decision behavior in single-exposure deterministic decision-making.

Neuropsychological evidence points to the MTL as a key substrate of deterministic learning and decision-making. Damage to the MTL routinely impairs ability on tasks such as discrimination learning (Buffalo et al., 1999; Corkin, 2002; Hood et al., 1999; Squire et al., 1988; Squire and Zola, 1996; Zola-Morgan et al., 1989; Zola-Morgan et al., 1994). However, in some cases, learning behavior can be spared at the cost of slower initial learning (Bayley et al., 2005;

Buffalo et al., 1998; Chudasama et al., 2008; Gaffan & Murray, 1992; Hood et al., 1999; Malamut et al., 1984; Phillips et al., 1988; Suzuki et al., 1993). Additionally, when examining extended learning trajectories, contributions of the MTL system tends to be overshadowed by the contributions of other systems, especially those of the striatum (Tremel et al., 2016, 2018). Thus, on the surface, it appears that the MTL might support critical processes early in a learning trajectory that use these initial experiences to build a mnemonic scaffold in preparation for extended learning. As learning improves through repeated experience, this initial scaffold may be supplanted by contributions from other slower systems, such as that mediated by the striatum, leading to the development of automatized, habitual behavior (Bayley et al., 2005; Packard & Knowlton, 2002; Seger & Cincotta, 2005; Tremel et al., 2016, 2018; Wimmer et al., 2012). This putative shift in systems may explain why any preserved learning in the absence of the MTL is generally slow and why striatal contributions tend to dominant in studies of extended learning (Bayley et al., 2005; Tremel et al., 2016, 2018).

The computational distinction between model-based and model-free reinforcement learning resembles this functional dichotomy between MTL and striatal contributions to decision-making (Gabrieli, 1998; Knowlton et al., 1996; Squire, 1992). In model-based reinforcement learning, decision-making is a prospective process wherein a model of the task environment is assembled and used to simulate potential outcomes of subsequent choices. Decision-making in this approach is driven by a policy derived from this model of the environment (Bornstein & Daw, 2013; Daw et al., 2011; Doll et al., 2012; Gillan et al., 2015; Gläscher et al., 2010). Activity in the MTL has been linked to prospective learning similar to that of this approach, reflecting the encoding of sequential relationships between transitions in a task (Doll et al., 2012; Doll et al., 2015a; Miller et al., 2017; Pennartz et al., 2011). Thus, model-

based reinforcement learning may be a viable mechanism underlying MTL contributions to decision-making.

In contrast to model-based learning, model-free reinforcement learning describes a retrospective process, wherein decision behavior is driven by a history of outcome value predictions that are refined through repetitive experience. In this process, the task environment is essentially ignored and instead, individual choice options accrue value through repeated experience. The striatum has been implicated as the foundation of a model-free reinforcement learning system, requiring many repetitions to build up accurate predictions of decision values that can inform behavior (Doll et al., 2015b; Tremel et al., 2016). Thus, in opposition to a model-based system that may support rapid learning, model-free systems are typically associated with longer, procedural learning trajectories (Gläscher et al., 2010; Niv, 2009; O'Doherty et al., 2015). Together, engagement of both model-based and model-free systems may underlie the ability to learn quickly from single experiences and subsequently develop that learning into automatized responses.

In the present study, I tested the possibility that the MTL is responsible for supporting a model-based learning system and that this computational mechanism facilitates initial decision-making after just one episode of experience. Subjects participated in a single-exposure deterministic learning task while in an fMRI scanner. Subjects learned to associate words with positive or negative reinforcement in a set of 80 word pairs in an initial exposure and then used that experience in a second exposure to make better decisions. Model-based and model-free reinforcement learning agents were constructed to learn this task based on the experiences of the subjects. Simulated decision outcome values derived from each of these models was then fit to regions in the MTL and the striatum to determine whether regional activity better reflected one

computational mechanism versus the other. I also characterized the extent to which individual subjects engaged in one learning approach versus the other to determine the behavioral consequences of preferring model-based versus model-free learning.

If the MTL is responsible for assembling an initial model or scaffold of the decision environment, its activity should predict model-based measures better than model-free measures. It was also expected that activity in striatal regions would better reflect model-free measures. This distinction would support the hypothesis that the MTL and striatal systems play different functional and computational roles in decision-making. If this distinction is responsible for initial decision successes after a single exposure, it was expected that individual differences in preference for model-based approaches would predict more accurate decision-making behavior and better memory for individual items. Thus, if model-based computation in the MTL can explain initial decision success, there should be performance benefits for relying more on that approach and system. Altogether, this study examines the hypothesis that model-based reinforcement learning is implemented by the MTL and that this MTL system facilitates decision-making after just one experience with a particular stimulus. This putative role could reconcile findings that MTL damage impairs learning with those implicating striatal systems as essential for deterministic learning.

3.1 MATERIALS AND METHODS

3.1.1 Subjects and task

The same data described in Chapter 2 were used here to investigate the computational mechanisms underlying deterministic decision-making. Briefly, 18 subjects (10 female, 20-25 years old, M = 21.56, SD = 1.79) participated in a deterministic decision-making task during a 2-hour behavioral and fMRI session. Subjects made decisions about 80 pairs of words and learned to associate their choices with positive or negative feedback. This first exposure to the word pairs and outcomes was trial-and-error (Round 1). They then experienced these word pairs a second time and used what they learned from Round 1 to make better decisions (Round 2). In Round 2, words selected in Round 1 were paired with a new word that had not yet been seen to account for memory differences due to selection (Tremel et al., 2018). Two factors were manipulated during this task. First, word pairs in Round 1 were presented on top of a background image consisting of either a clear, detailed natural landscape or a noise-degraded landscape. Second, word pairs were associated with either low ($0.10 to $0.29) or high ($0.81 to $1.00) monetary values. Feedback following each choice in Round 1 indicated specific magnitudes of the monetary gains or losses. These factors were manipulated to specifically target the engagement of the MTL and the striatum, respectively.

Following the deterministic decision-making task and scan session, subjects participated in a surprise memory test. Words from the deterministic decision-making task and words drawn from a new list were presented individually as subjects responded to two memory probes. First, subjects reported whether they recognized the word from the scan session or not (recognition memory). Decision task words that were correctly recognized were considered “hits” (versus

"misses"), while new words that were mistakenly reported as recognized were considered "false alarms" (versus "correct rejections"). Second, subjects reported whether they remembering experiencing that word with positive (gain) or negative (loss) feedback in Round 1 (episodic memory). Overall, this task assessed the state of subjects' memories after engaging in deterministic decision-making.

3.1.2 Reinforcement learning agents

Given the presence of both model-free and model-based computation in the brain, it is possible that one approach dominates in deterministic decision-making. The MTL in particular may engage in model-based learning that is sufficient to drive behavior without the influence of model-free signals from regions such as the striatum. As a consequence, it is unclear how model-based and model-free signals each contribute to deterministic decision-making and where the MTL and striatum fit within that scope. To assess these competing accounts about how the MTL and striatum learn from deterministic experiences, three types of reinforcement learning agents were implemented. Two models captured different aspects of learning based on either an accrued history of experience or predictions from a model of the task environment. A third approach combined these two styles in a hybrid model to assess subject-level preference for engaging in one computation versus the other.

3.1.2.1 Model-free agent

To capture the dynamics of model-free, retrospective learning, an off-policy temporal difference control algorithm (Q-learning) was implemented. In Q-learning, an agent updates predictions about quality values, *Q*, for particular decisions by making a choice and assessing the observed

reward. In the present task, each word pair is represented by a state, *s*, and each possible choice is represented by an action, *a*. Thus, each word in a pair is associated with a state-action quality value, *Q (s, a)*. Because subjects have no prior experience with the word pairs but may have a non-zero and non-systematic propensity to choose one item over another, *Q* values were instantiated as a small random value between -0.1 and 0.1, where 0 indicates a choice with no value, +1 indicates a rewarding choice, and -1 indicates a punishing choice. As the agent engages in Round 1 of the task, it updates each *Q (s, a)* in accordance with a subject's behavior based on a reward prediction error, δ_{RPE}. δ_{RPE} captures the difference between an observed outcome and predicted outcome, computed as:

Equation 1. Model-free reward prediction error.

$$\delta_{RPE} = r(s') - Q(s,a)$$

where *r(s')* is the observed reward for the selected action and *Q(s,a)* is the previous quality value associated with that state-action pair. This prediction error is used to update state-action values between Round 1 and Round 2 as:

Equation 2. Q-learning state-action value update

$$Q(s,a) = Q(s,a) + \alpha \times \delta_{RPE}$$

where α is a free parameter that controls the learning rate of the agent, or the extent to which observed information impacts the updating of prior beliefs. *Q(s,a)* was not updated for Round 2 choices since subjects received no explicit feedback about the outcomes. Likewise, *Q(s,a)* was not updated for unselected items since selected items in Round 1 are re-paired with new items in

Round 2. Optimal agents should prefer actions associated with higher state-action values than actions with lower values.

3.1.2.2 Model-based agent

Model-based agents represented the deterministic decision-making task as a matrix of transition probabilities between two given states. These agents updated this matrix of transition probabilities, *T (s, a, s')*, based on subjects' experience with Round 1. Because this task is deterministic and all observed probabilities are either 0 or 1, the state space for *s'* was reduced to two states. Thus, Round 1 pairs (*s*) and actions (*a*) were each represented as a state-action pair similar to that of model-free reinforcement learning, but Round 2 states (*s'*) represented which action would lead to a correct decision. In other words, *s'* reflected whether a subject should reselect the same word in Round 2 that was selected in Round 1, or whether the Round 1 word should be avoided and the other word selected. Thus, correct selections in Round 1 should lead to a Round 2 state in which the same selection should be made. Errors in Round 1 should lead to Round 2 states in which subjects should select the new word. The matrix of transition probabilities therefore holds an estimate of the probability that a given action in Round 1 will transition to a particular state in Round 2. Transition probabilities were instantiated at 0.5, assuming that subjects were unbiased with respect to which action they choose during Round 1.

Agents update this transition probability matrix via a state prediction error, δ_{SPE}, computed as:

Equation 3. Model-based state prediction error for Round 1 gains.

$$\delta_{SPE}(gain) = 1 - T(s, a, s')$$

for correct trials (i.e., gains) in Round 1 and as:

Equation 4. Model-based state prediction error for Round 1 losses.

$$\delta_{SPE}(loss) = 0 - T(s, a, s')$$

for error trials (i.e., losses) in Round 1. A transition probability is updated based on the observed transition from *s* to *s'* via:

Equation 5. Model-based transition probability update.

$$T(s, a, s') = T(s, a, s') + \eta \times \delta_{SPE}$$

where η is a free parameter controlling the model-based learning rate, or the extent to which new information about state transitions impacts the update of the current probability matrix. To ensure that the distribution of probabilities remains normalized, probabilities for states that are never visited are reduced via:

Equation 6. Probability update for unvisited "gain" states in model-based learning.

$$T(s, a, s') = T(s, a, s') \times (1 - \eta)$$

for positive reinforcement (reward) trials, and via:

Equation 7. Probability update for unvisited "loss" states in model-based learning.

$$T(s, a, s') = T(s, a, s') \times (1 + \eta)$$

for loss trials. State transition probabilities associated with the untaken action remained unchanged since the subject received no information about that action. Effectively, the agent does not make use of these unchanged values since the computation of state-action values, *Q*, depends on having taken a particular action in Round 1. Appendix E illustrates this process using two example trials. These estimated transition probabilities were then used to compute a state-action value for each Round 2 state-action pair, *Q (s, a)*, by:

Equation 8. Computation of state-action values in model-based learning.

$$Q(s,a) = T(s,a,s') \times r(s)$$

where *r* is the observed reward received by subjects for that state-action pair. Thus, model-based agents first assemble a model of the task environment based on experience and subsequently use this model to compute state-action values to drive their decision-making.

3.1.2.3 Hybrid agent

It is likely that individuals engage in multiple styles of reinforcement learning simultaneously. Gläscher and colleagues (2010) demonstrated that model-based learning is feasible in humans and seems to be implemented in various brain regions. Similarly, Daw et al. (2011) showed evidence that different sub-sections of the striatum in particular may engage preferentially in either model-free or model-based reinforcement learning. Thus, in line with these prior studies, this third possibility was considered, wherein some brain regions may engage in both styles of learning but may prefer one approach to the other. To assess this, a hybrid reinforcement learning agent was implemented utilizing the learning rate parameters from each of the model-free and model based agents. State-action values were computed using the model-free and

model-based state-action values obtained from simulations with optimal learning rate parameters. Hybrid state-action values, $Q_{HY}(s, a)$, were computed as a weighted sum:

Equation 9. Computation of state-action values for the hybrid agent.

$$Q(s,a)_{HY} = w \times Q(s,a)_{MB} + (1-w) \times Q(s,a)_{MF}$$

where Q_{MB} and Q_{MF} are state-action values from the model-based and model-free agents, respectively, and w is a free parameter representing the relative weight of engagement with model-based learning versus model-free learning. Values of w closer to 1.0 indicate a preference for model-based reinforcement learning, while values approaching 0.0 indicate a preference for model-free learning.

3.1.2.4 Optimization of agents

Agents were optimized assuming that subjects behave stochastically based on probabilities derived from the distribution of state-action values. Choice probability estimates were computed using a softmax action selection rule:

Equation 10. Choice probability estimation from softmax function.

$$P(s,a) = \frac{e^{\tau \times Q(s,a)}}{\sum_{b=1}^{n} e^{\tau \times Q(s,b)}}$$

where Q is the state-action value derived from an agent, and τ is the inverse temperature, controlling the extent to which an agent prefers the option with the highest value. Each type of agent had two free parameters, including learning rate for model-free (α) and model-based (η)

agents, the weighting parameter (w) of the hybrid agent, and the inverse temperature of the softmax selection rule (τ). Optimal free parameters for each subject's agents were recovered by minimizing the negative log likelihood based on the probabilities derived from the softmax function:

Equation 11. Likelihood function.

$$-LL = -\sum \log(\ P\ (s, a)\ )$$

For each subject, multiple Nelder-Mead minimizations (n = 100) were run to recover optimal free parameter values. Parameters from the top 10% of these minimizations (based on negative log-likelihood) were averaged to ensure that the recovered parameters were adequately representative of each subject's behavior. These average parameters were considered optimal. Simulations (n = 1000) were then run for each agent using the optimal free parameter values to obtain trial-level estimates of state-action value and prediction error metrics. The simulation associated with the lowest negative log likelihood was used for analysis. Goodness-of-fit was compared within each subject using likelihood ratio tests and AIC differences compared to a null model. Null models were chance-performance agents wherein the response probabilities were 0.5 for all trials.

Model performance was assessed by computing the trial-level accuracy score for each agent (i.e., whether the agent's selection on each trial correct or incorrect in terms of the task) and compared to the corresponding subject's behavior. If the agent performed identically to a subject on a trial, that trial was considered a match. The total percentage of matches between agent behavior and subject behavior was used to quantify the extent of fit.

3.1.3 Image processing, region identification, and time series extraction

Imaging data were identical to that analyzed in chapter 2. Briefly, key regions in the MTL and the striatum were identified to test the extent to which each region engages in different types of reinforcement learning. Briefly, these regions were identified using a secondary dataset that used a concurrent discrimination learning task which featured a similar deterministic reinforcement schedule (Tremel et al., 2016; Tremel et al., 2018). A repeated-measures ANOVA was computed on the first two rounds of concurrent discrimination learning to localize voxels that were active during deterministic learning (i.e., exhibited a main effect of time), restricted to MTL and striatal territory. The resulting statistical map was corrected for multiple comparisons and sphericity. Clusters of voxels surviving this correction were considered regions of interest. These regions are reported in Table 2. Activity time series were extracted by computing a null-effect GLM and computing the model residuals. The residual time series was expressed as a percent change from the model baseline term. Trial-level time series were assembled by segmenting the full residual time series into 11 time point sequences starting from the onset of each trial.

3.1.4 Functional region preference for model-based or model-free reinforcement learning

The first goal of this study was to link activity in the MTL and the striatum to model-based and model-free reinforcement learning computation. Mixed-effects linear regression was used to assess the extent to which each region of interest in the MTL and striatum engaged in model-free versus model-based learning. To test this, two regression models were computed per region using the agent-derived state-action choice values as a predictor of regional activity (i.e., fMRI activity). One regression was computed for the model-free measures and one for the model-based

measures. The weight parameter, *w*, derived from the hybrid reinforcement agent was also entered as a subject-level predictor of regional activity and was allowed to interact with the agent-derived predictor. Interactions between state-action values and the *w* parameter would indicate that a region's activity is sensitive to the relative engagement of model-free and model-based approaches at a subject level. For example, subjects who are more inclined to engage in a model-free approach (i.e., lower *w* value) may exhibit greater activity in some regions versus others, signifying that those regions are more engaged in model-free learning versus model-based learning. In contrast, if a region does not exhibit an interaction between state-action values and the *w* parameter, that region would be engaged in either model-free or model-based learning regardless of preference. Because the weight parameter was included as a predictor in both models to account for the relative engagement in model-free or model-based learning, state-action values from the hybrid agents were not analyzed. To remain consistent with and facilitate comparisons to prior work using reinforcement learning agents, average activity from time points 4-7 was used as the dependent variable (Tremel et al., 2016). Best-fitting regression models were selected based on AIC and likelihood ratio tests. Typically, differences in AIC greater than 2.0 are considered sufficient evidence to favor one model over another (Burnham & Anderson, 2003).

It was predicted that, if the MTL is critical for scaffolding early learning, activity in the MTL should relate to model-based computation and subjects should generally show preference to model-based learning (i.e., high *w* parameter). On the other hand, if the striatum is the primary driver of this type of learning, subjects should generally show preference to model-free learning (i.e., low *w* parameter). Realistically, it was expected that subjects would engage in both model-

free and model-based learning, but the relative preference would favor model-based computation in support of the idea that the MTL builds a scaffold from initial encoding experiences.

In a secondary analysis, the preferential engagement of particular regions in model-free or model-based learning was characterized at a between-subjects level. To do this, the average regional activation of each region during Round 2 decision-making of three sub-groups of subjects was examined. Since very few subjects exhibited a preference for model-free learning ($w < 0.5$, N = 4), subjects were split into three groups using arbitrary cut-off points. The first group of subjects exhibited a slight preference for model-free learning based on the w parameter ($w < 0.50$). An equally sized group (N = 4) of subjects was created using a range from $w = 0.50$ to 0.60. This group exhibited a slight preference for model-based learning. All other subjects exhibited a relatively strong preference for model-based learning, with a w value greater than 0.60. Average regional activity was compared across groups for each region using between-subjects ANOVAs. While this analysis may be underpowered due to small groups relative to the regression reported above, it offers a glimpse of the general trends of which regions are important for individuals engaging in different learning approaches.

3.1.5 Behavioral profile of model-free and model-based learning

The second goal of this study was to establish a behavioral profile associated with engaging in model-free or model-based learning. It was expected that, if the MTL were important for building a scaffold from early decision experiences, subjects exhibiting preference for model-based learning would also exhibit superior decision-making accuracy in Round 2 and higher quality memory for the learned items. To assess the behavioral consequences of engaging in a model-free or model-based approach, logistic regression was used to test the relationship

between the *w* parameter and accuracy measures for decision-making and subsequent memory. The *w* parameter was entered as a subject-level predictor in three regression models of Round 2 decision-making accuracy, recognition memory accuracy, and episodic memory accuracy. Round 1 outcome (gain or loss) was also included as a predictor since initial outcome valence can impact decision and memory abilities (Tremel et al., 2016). These two predictors were allowed to interact in the models, since initial outcome valence could differentially affect accuracy for individuals preferring one reinforcement learning approach to another. Subject was included as a random effect.

3.2 RESULTS

3.2.1 Reinforcement learning agent parameters and fits

In this study, model-based and model-free reinforcement learning agents were fit to a single-exposure deterministic learning task to assess the relative impact of each on decision-making processes. In particular, it was hypothesized that the MTL and model-based approaches would be critical in supporting initial decision-making, while model-free learning via the striatum may be supplementary. After fitting the agents to the data, optimal parameters were used to simulate decision outcome values that were in turn used to predict fMRI activity in the MTL and striatum and behavioral measures of decision-making and memory.

Each agent had a learning rate free parameter, which controls the extent to which new experiences impacts the updating of previous memory (Table 7). For the model-free agent, the mean learning rate, α, was 0.41 (SD = 0.35). The mean model-based learning rate, η, was 0.89

(SD = 0.06), suggesting that on average, model-based agents produced better fits at higher learning rates than the model-free agents. While the hybrid agent utilizes learning rates from both model-free and model-based agents, it had an additional free parameter, *w*, which controlled the relative balance of model-based versus model-free computation. Weight values above 0.5 signified a preference for model-based learning, while values below 0.5 signified preference for model-free learning. On average, subjects preferred model-based approaches (M = 0.64, SD = 0.13), which suggests that model-based learning is capable of producing single-exposure deterministic decision-making behavior.

Table 7. Reinforcement agent free parameters

Subject	α	η	*w*
1	0.76	0.85	0.63
2	0.92	0.83	0.93
3	0.57	0.94	0.59
4	0.15	0.94	0.56
5	0.35	0.85	0.67
6	0.13	0.94	0.57
7	0.53	0.81	0.78
8	0.93	0.87	0.69
9	0.06	0.90	0.50
10	0.08	0.94	0.44
11	0.92	0.73	0.69
12	0.66	0.87	0.71
13	0.19	0.94	0.46
14	0.04	0.94	0.72
15	0.14	0.94	0.55
16	0.09	0.93	0.48
17	0.79	0.82	0.85
18	0.06	0.95	0.61

Three separate reinforcement learning agents were fit using each subject's behavioral data. α, model-free agent learning rate; η, model-based agent learning rate; *w*, weight parameter of the hybrid agent.

In general, all three agents performed well in terms of goodness-of-fit compared to the null model of random choice probabilities (Table 8). Each agent was also able to match trial-level subject behavior at rates above 80% on average, with model-free learning at 83% (SE = 4%), model-based learning at 82% (SE = 3%), and the hybrid agent at 89% (SE = 2%). The hybrid learner exhibited the highest percentage, indicating that behavior is best represented by a weighted combination of model-free and model-based approaches.

Table 8. Reinforcement agent fit statistics

	Model-free				*Model-based*				*Hybrid*			
Subject	LL	AIC	ΔAIC	Match	LL	AIC	ΔAIC	Match	LL	AIC	ΔAIC	Match
1	916.0	1834.1	54.53	0.99	913.0	1830.1	70.67	0.92	915.6	1833.3	56.64	0.97
2	911.6	1825.1	63.45	1.00	910.8	1825.5	75.23	0.95	913.7	1829.3	60.60	0.97
3	920.8	1843.5	45.03	0.95	914.4	1832.7	68.00	0.80	915.5	1832.9	56.99	0.95
4	921.7	1845.4	43.17	0.74	915.1	1834.2	66.56	0.74	915.9	1833.9	56.05	0.90
5	920.4	1842.7	45.83	0.78	914.8	1833.7	67.06	0.82	916.3	1834.7	55.22	0.91
6	921.7	1845.3	43.21	0.76	915.0	1834.0	66.75	0.75	915.8	1833.6	56.29	0.86
7	919.3	1840.6	47.85	0.95	914.1	1832.3	68.47	0.94	916.0	1833.9	55.99	0.99
8	912.7	1827.3	61.26	0.99	911.4	1826.9	73.86	0.95	914.4	1830.9	59.04	0.97
9	921.8	1845.6	42.95	0.61	917.0	1837.9	62.84	0.66	917.2	1836.3	53.59	0.69
10	921.9	1845.8	42.79	0.70	917.1	1838.2	62.49	0.75	917.8	1837.7	52.21	0.79
11	913.7	1829.3	59.23	0.99	912.0	1828.0	72.76	0.92	914.8	1831.7	58.22	0.97
12	917.9	1837.8	50.76	0.99	913.2	1830.4	70.31	0.94	915.1	1832.3	57.63	0.96
13	921.6	1845.3	43.28	0.80	915.3	1834.6	66.19	0.76	916.1	1834.3	55.65	0.87
14	921.8	1845.6	43.00	0.60	916.9	1837.6	63.09	0.77	916.8	1835.6	54.36	0.77
15	921.1	1844.2	44.34	0.86	915.5	1835.0	65.75	0.81	916.3	1834.6	55.33	0.94
16	922.1	1846.1	42.46	0.75	917.1	1838.3	62.47	0.61	917.8	1837.7	52.26	0.79
17	916.0	1834.0	54.54	0.98	912.8	1829.6	71.11	0.92	915.7	1833.4	56.51	0.99
18	922.0	1846.0	42.58	0.51	917.9	1839.9	60.85	0.64	918.1	1838.2	51.72	0.67

Fit statistics were computed separately for each subject and for each of the three types of reinforcement learning agents. *LL*, negative log-likelihood of the model; *AIC*, Akaike's Information Criterion; *ΔAIC*, difference in AIC between the model and a null model fit to a random chance-performance trial sequence; *Match*, proportion of total trials for which the agent's choice matched the choice of the subject (1.0 indicates a perfect match between agent behavior and human behavior).

3.2.2 Preference of the MTL and striatum for model-free versus model-based computation

The primary goal of this study was to establish a connection from decision-related activity in the MTL and striatum to underlying reinforcement learning computational mechanisms. It was expected that the MTL would be more associated with model-based computation, whereas the striatum would be more associated with model-free computation. Moreover, to test the hypothesis that the MTL may be involved in building a memory scaffold from early decision experiences, the extent to which these regional computation differences modulated in accordance with a weight parameter from a hybrid reinforcement learning model was assessed. In this primary analysis, mixed-effects regression models were computed to test the extent to which model-free or model-based state-action values, $Q(s, a)$, could predict activity of each region. The w parameter was included as a predictor to account for subject-level preference for one approach to the other and was allowed to interact with the state-action value predictor. Two regions in the caudate nucleus (body and head) exhibited preference for model-free computation. Three regions, including the hippocampus, the bilateral parahippocampal gyrus, and the left parahippocampal gyrus exhibited preference for model-based computation. The remaining two regions, including the left hippocampus and left putamen exhibited no correlation with either measure and were excluded from further analysis. Model comparison and fit statistics of these regions are enumerated in Table 9.

Table 9. Regression fits

	Model-free		*Model-based*				
Region	LL	AIC	LL	AIC	ΔAIC	p	Best agent
R Hippocampus	2742.5	5497.1	2730.2	5472.3	24.8	<0.001	Model-based
L Hippocampus	1571.3	3154.6	1570.4	3152.8	1.8	1.00	--
Bilat. Parahippocampal G	453.1	918.1	397.8	807.6	110.5	<0.001	Model-based
L Parahippocampal G	2835.8	5683.7	2827.3	5666.5	17.2	<0.001	Model-based
L Body of Caudate	1420.8	2853.7	1427.8	2867.6	-13.9	<0.001	Model-free
Bilat. Head of Caudate	156.3	300.6	160.9	309.7	-9.1	<0.001	Model-free
L Putamen	1111.5	2235.0	1111.6	2235.1	0.1	1.00	--

Regression models were computed to test whether activity in each region of interest better related to choice quality values derived from model-based or model-free agents. *LL*, log-likelihood of the regression model; *AIC*, Akaike's Information Criterion of the regression model; *ΔAIC*, difference in AIC between the model-free regression and the model-based regression (negative values indicate better model-free fit); *p*, p-value of the likelihood ratio test comparing the two models; *Best agent*, the type of agent that best fit activity in the region; *Bilat*, bilateral; *R*, right; *L*, left; *G*, gyrus.

Activity in three regions modulated with model-based state-action values, including the right hippocampus, the bilateral parahippocampal gyrus, and the left dorsal parahippocampal gyrus. Activity in all three regions negatively correlated with state-action values such that positive activity coded for negative value predictions and negative activity coded for positive value predictions (Figure 11). Additionally, there was an interaction between value predictions and the *w* parameter in the bilateral parahippocampal gyrus wherein activity better predicted state-action values at higher weights than lower weights. At higher weights, there was a negative relationship between value predictions and activity, however at lower weights, this relationship was absent (Figure 11).

Table 10. Regression statistics

Region	*Effect*	*Coef. β*	*SE β*	*t*	*p*		*Best agent*
R Hippocampus	*w*	0.90	0.53	1.69	0.11		Model-based
	Q (s, a)	-0.21	0.04	-5.30	<0.001	*	
	w × *Q (s, a)*	0.55	0.30	1.82	0.07		
Bilat. Parahippocampal G.	*w*	-0.06	0.17	-0.38	0.70		Model-based
	Q (s, a)	-0.20	0.02	-11.73	<0.001	*	
	w × *Q (s, a)*	-0.49	0.13	-3.81	<0.001	*	
L Parahippocampal G.	*w*	0.75	0.87	1.30	0.21		Model-based
	Q (s, a)	-0.18	0.04	-4.41	<0.001	*	
	w × *Q (s, a)*	-0.01	0.31	-0.04	0.97		
L Body of caudate	*w*	-0.17	0.27	-0.64	0.53		Model-free
	Q (s, a)	0.13	0.05	2.34	0.02	*	
	w × *Q (s, a)*	-0.76	0.35	-2.17	0.03	*	
Bilat. Head of caudate	*w*	-0.01	0.18	-0.08	0.94		Model-free
	Q (s, a)	0.07	0.03	2.12	0.02	*	
	w × *Q (s, a)*	-0.19	0.20	-0.97	0.33		

Regression models were computed for each region to test the extent to which regional activity could be predicted by choice-values from a model-based or model-free reinforcement learning agent (indicated by *Best agent*) and by the *w* weighting parameter from the hybrid model. Coef. *β*, regression coefficient; SE, standard error of coefficient; *t*, t-statistic of effect; *p*, p-value of effect; asterisk indicates statistically significant effect.

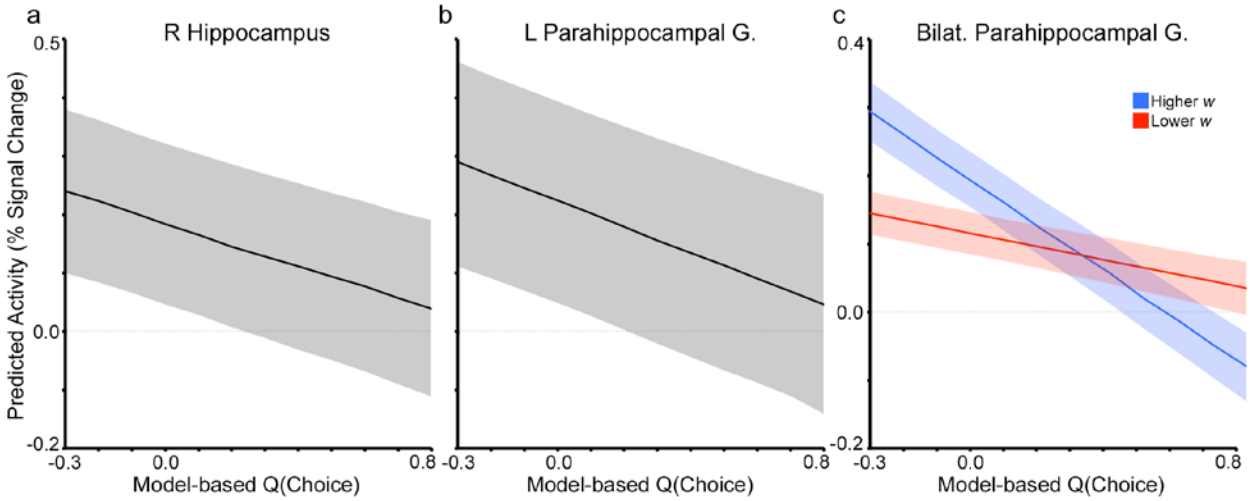

Figure 11. Model-based computation in the MTL.

Model-based quality value estimates were predictive of activity in three regions of the MTL. (a) Predicted activity in the right hippocampus plotted against choice quality values from the model-based agent. (b) Predicted activity in the left parahippocampal gyrus versus model-based choice quality values. (c) Interaction effect in the bilateral parahippocampal gyrus wherein activity was better predicted by model-based choice quality values for individuals with a greater preference for model-based versus model-free computation (higher *w* parameter). Data for the *higher w* group is from six subjects with the highest *w* parameter values, whereas data for the *lower w* group is from six subjects with the lowest *w* parameter values. Shaded areas reflect standard error of the mean.

Activity in the two model-free regions, the left body of the caudate nucleus and the bilateral head of the caudate nucleus, was positively correlated with the magnitude of model-free state-action choice values (Table 10). Additionally, there was an interaction between state-action values and the weight parameter, *w*, in the caudate body, indicating that the extent to which model-free value predictions correlated with activity depended on individual preference for model-free learning (Figure 12). Activity positively correlated with state-action values at lower weights (i.e., greater model-free preference), but was anti-correlated for higher weights.

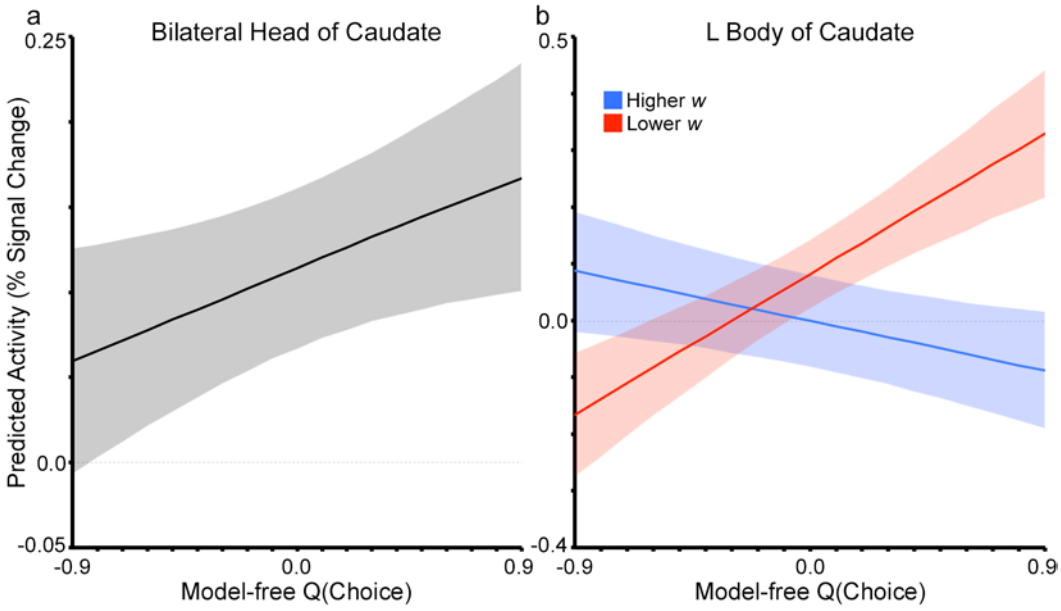

Figure 12. Model-free computations in the striatum.

Model-free quality value estimates were predictive of activity in two regions of the striatum. (a) Predicted activity in the bilateral head of the caudate nucleus plotted against choice quality values from the model-free agent. (b) Interaction effect in the left body of the caudate nucleus wherein activity was better predicted by model-free choice quality values for individual with a greater preference for model-free versus model-based computation (lower *w* parameter). Data for the *higher w* group is from six subjects with the highest *w* parameter values, whereas data for the *lower w* group is from six subjects with the lowest *w* parameter values. Shaded areas reflect standard error of the mean.

In a secondary analysis, mean regional activity was computed for three sub-groups corresponding to the extent of preference for model-based learning. Ten subjects had a *w* value

greater than 0.6, suggesting a strong preference for model-based learning. Four subjects fell in a range between 0.5 and 0.6, indicating a slight preference for model-based learning. The four remaining subjects had a *w* value under 0.5, indicating a slight preference for model-free learning. A between-subjects ANOVA was computed for each region, testing the magnitude of activation between the groups. One region in the right hippocampus exhibited differences across groups such that activity was greatest for the strong model-based preference group, approximately zero for the slight model-based preference group, and negative for the slight model-free preference group, $F[2, 15] = 4.25$, $p = 0.03$ (Figure 13). Other regions, however, exhibited no statistically significant effects, including the left caudate body, $F[2, 15] = 0.25$, $p = 0.78$, the bilateral caudate head, $F[2, 15] = 0.23$, $p = 0.80$, the bilateral parahippocampal gyrus, $F[2, 15] = 0.10$, $p = 0.91$, and the left parahippocampal gyrus, $F[2, 15] = 1.36$, $p = 0.29$.

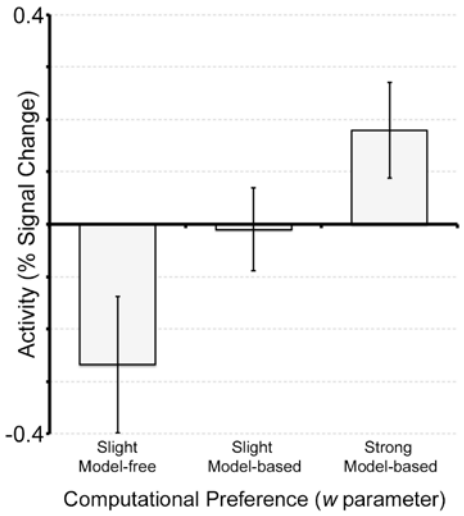

Figure 13. Preferential recruitment of the hippocampus for model-based learners

Mean activity in the right hippocampus (as % signal change from baseline) is plotted for three sub-groups of subjects based on their preference for model-free or model-based learning (*w* parameter value). Slight model-free preference (N = 4) is defined as $w < 0.5$; slight model-based preference is defined as $0.5 < w < 0.6$; strong model-based preference is defined as $w > 0.6$.

3.2.3 Behavioral profile of model-free and model-based learning

The second goal of this study was to characterize the consequences of preferentially engaging in model-free or model-based learning. Given the association between the MTL and model-based computation and that subjects tend to prefer model-based learning in this task, it was expected that subjects preferring a model-based approach would make better decisions in Round 2 and would retain stronger memories of individual items. This would provide additional evidence that the MTL is essential to supporting an early memory foundation that can then be used to drive initial decision-making while model-free learning systems can work in parallel to develop automaticity through repetition. To probe the behavioral profiles associated with model-free and model-based learning, logistic regressions were computed testing the relationship between the *w* parameter, Round 1 outcome (i.e., initially receiving a gain a loss), and accuracy measures of decision-making and subsequent memory.

Successful decision-making in Round 2 was found to be dependent on the extent to which subjects prefer model-based learning (*w* predictor), $\beta = 6.92$, $SE_\beta = 1.28$, $z = 5.41$, $p < 0.001$. Subjects engaging in model-free learning generally exhibited poorer decision-making in Round 2. As preference for model-based learning increased, so did decision-making accuracy (Figure 14). There was also an effect of Round 1 outcome, wherein decision accuracy in Round 2 was overall higher for items experienced with positive (gain) feedback in Round 1 relative to items experienced with negative (loss) feedback in Round 1, $\beta = 0.90$, $SE_\beta = 0.66$, $z = 3.31$, $p < 0.001$, but no interaction between the two predictors, $\beta = -2.84$, $SE_\beta = 1.08$, $z = 0.89$, $p < 0.37$).

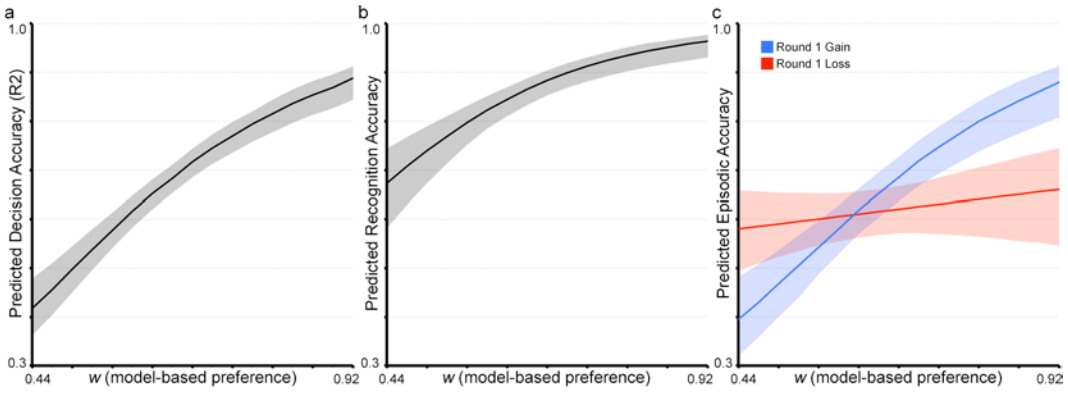

Figure 14. Behavioral consequences of the model-based versus model-free tradeoff

Individual preference for model-based learning was predictive of behavioral decision and memory outcomes. (a) Probability of making a correct response during Round 2 of the deterministic decision-making task increased in proportion to individual preference for model-based learning. (b) Probability of correctly recognizing an item after the task in a surprise memory test increased in proportion to individual preference for model-based learning. (c) Probability of correctly recalling details about individual items in a surprise memory test increased in proportion to individual preference for model-based learning, but in a manner that depended on the whether that item was associated with a gain or loss in the decision task. Shaded areas reflect standard error of the mean.

Recognition memory ability also depended on preference for model-based learning (*w* predictor), $\beta = 5.24$, $SE_\beta = 1.87$, $z = 2.81$, $p < 0.01$ (Figure 14). However, there was no effect of Round 1 outcome, $\beta = -0.38$, $SE_\beta = 0.79$, $z = -0.48$, $p = 0.63$, and no interaction between the predictors, $\beta = 0.38$, $SE_\beta = 1.31$, $z = 0.29$, $p = 0.77$. This indicates that the ability to recognize items after the decision-making task was generally improved by relying more on a model-based approach versus a model-free approach.

Episodic memory accuracy was additionally related to model-based preference, $\beta = 5.05$, $SE_\beta = 1.62$, $z = 3.12$, $p < 0.01$, wherein accuracy increased as preference for model-based learning increased (Figure 14). While there as no effect of Round 1 outcome, $\beta = 2.69$, $SE_\beta = 0.58$, $z = 0.19$, $p = 0.85$, there was an interaction between outcome and *w*, $\beta = -4.37$, $SE_\beta = 0.92$, $z = -4.78$, $p < 0.001$. For items experienced as gains in Round 1, the *w* parameter was predictive of episodic memory accuracy, but this was not the case for items experienced as losses in Round 1.

3.3 DISCUSSION

3.3.1 Model-based reinforcement learning via the MTL supports single-exposure decision-making

In the single-exposure decision experiences of the present task, individuals who preferred model-based approaches were behaviorally more successful than those who favored more model-free involvement. This indicates that model-based approaches may be necessary to support rapid learning from an initial exposure. Single-exposure decision-making may rely on the ability to successful construct a model of the environment rapidly and use that model to predict future outcomes.

Individuals with the strongest preference for model-based approaches also exhibited the strongest declarative memory for individual items later. Successful declarative memory recollection may then require the ability to plan for future outcomes based on a model developed from initial experiences (Hassabis & Maguire, 2009; Schacter et al., 2012). This may partly explain why impairments to declarative memory can drastically reduce decision-making ability after an initial learning experience (Buffalo et al., 1999; Corkin, 2002; Hood et al., 1999; Squire et al., 1988; Squire and Zola, 1996; Zola-Morgan et al., 1989; Zola-Morgan et al., 1994). Consistent with this, individuals who least preferred model-based approaches performed the worst on measures of decision-making and subsequent memory. Thus, model-based reinforcement learning may capture the process by which an initial experience is translated into successful decision-making without additional repetition.

Given the association between declarative memory and model-based learning, it makes sense then that a neural substrate associated with declarative memory, the MTL system, was

associated with model-based learning. The MTL has been implicated in implementing model-based processes to support decision-making after an initial experience (Chalmers et al., 2016; Kaplan et al., 2017; Miller et al., 2017). In the present study activity in the MTL was better fit by model-based measures than by model-free ones, suggesting that the MTL may build representations of the environment from an initial exposure and use this model to make better decisions at a second exposure. Additionally, individuals who most preferred model based approaches exhibited greater engagement of the MTL than individuals who tended toward model-free approaches. This adds further evidence to the idea that an MTL declarative memory system may be critical for supporting decision-making behavior after just a single experience.

These findings suggest that the MTL may facilitate the building of a mnemonic scaffold of declarative memory that is used to support initial decision-making. This scaffold is crucial to initial learning success since the alternative model-free processes tend to require several repetitions to refine the veracity of their outcome value predictions (Daw & Doya, 2006; Dayan & Daw, 2008; O'Doherty et al., 2015; Yin et al., 2004, 2005). It is possible that after several repetitions when model-free predictions are more reliable, this scaffold becomes unnecessary and that effortful declarative recall gives way to habit-based recall supported by model-free learning (Seger & Spiering, 2011; Seger et al., 2010; Tremel et al., 2016; Tricomi et al., 2009). This may explain why prior work has failed to link MTL activity to decision-making behavior during extended learning (Tremel et al., 2016, 2018), wherein the striatum was found to dominate. If the MTL is critical for supporting the first few decision experiences, examining a long learning trajectory may be a poor method for detecting those contributions. Indeed, prior work that specifically manipulated the efficacy of declarative memory during extended learning was more

successful in detecting a shift toward model-free striatal engagement instead of detecting an enhanced model-based hippocampal engagement (Tremel et al., 2018).

It is worth noting that while model-based reinforcement learning has been implemented for probabilistic task structures, it is not usually implemented for a deterministic structure. This perspective is traditionally neglected because it is difficult to fit models when state transitions are associated with 0% or 100% probabilities (i.e., one decision will always lead to one state, instead of one decision favoring the transition to a particular state over another). Thus, while the approach of the present study may have weaknesses (e.g., the 80 pairs of Round 2 are reduced into two possible states), the findings are generally consistent prior findings, namely those that have implicated the hippocampus in model-based learning (Doll et al., 2015a; Duncan et al., 2018; Gershman & Daw, 2017; Miller et al., 2017). However, other findings have suggested that different sub-regions of the striatum may differentially implement model-based versus model-free learning (Daw et al., 2005, 2011). This distinction was not examined in the present study, but it is likely that this is also true for deterministic task structures and warrants further investigation.

3.3.2 Both model-based and model-free systems are engaged in parallel

While individuals in the present study generally exhibited a preference for model-based reinforcement learning, most individuals implemented a mix of model-based and model-free processes. Indeed, the hybrid reinforcement learning models were generally the best-fitting to behavior, wherein choices from the agent frequently matched choices from the human subjects. This suggests that both approaches are implemented in the brain and that decision behavior results from the integration of multiple types of information (Bornstein & Daw, 2013; Daw et al.,

2011; Doll et al., 2012; Doll et al., 2015b; Gläscher et al., 2010). Regions in the prefrontal cortex have been implicated in arbitrating between model-based versus model-free approaches and may also serve to integrate these multiple types of information (Doll et al., 2015b; Lee et al., 2014, 2014; Poldrack & Rodriguez, 2004).

Though both approaches seem to be implemented in the brain, model-based and model-free computations seem to be mediated by distinct neural systems. Model-based computation seems to be mediated preferentially by the MTL system, whereas model-free computation is better fit by activity in the striatum. Activity in the caudate coded for value prediction signals derived from an accrued history of choice selection. Additionally, individuals who tended to prefer more model-free approaches seemed to engage the caudate moreso than individuals who preferred model-based approaches. However, it is notable that model-free learning seems to be an unreliable approach in cases of single-exposure learning such as this since individuals preferring this approach performed worst on measures of subsequent decision-making and memory.

The present study found a dissociation between regions that were better fit by model-based processes (i.e., MTL regions) versus those better fit by model-free ones (i.e., striatal regions). Each of these neural systems is associated with different roles in learning and decision-making, wherein prospective events (e.g., expectancy) localize to the MTL and retrospective events (e.g., surprise) localize to the striatum (Bornstein & Daw, 2012; Gläscher et al., 2010). This neural and computational dichotomy can map onto an additional cognitive distinction between declarative and procedural processes (Gabrieli, 1998; Knowlton et al., 1996; Squire, 1992). As such, this study provides additional evidence that parallel processes in separate neural systems may underlie decision behavior. These systems may be engaged differentially depending

on contextual and structural factors of a task (Tremel et al., 2016, 2018), but also depending on individual differences and the amount of experience gained.

The parallel engagement of MTL and striatal systems via model-based and model-free reinforcement learning, respectively, may explain apparent discrepancies in neuropsychological and neuroimaging findings. MTL damage in humans and monkey produces profound learning deficits in deterministic learning tasks like concurrent discrimination (Buffalo et al., 1999; Corkin, 2002; Hood et al., 1999; Squire et al., 1988; Squire and Zola, 1996; Zola-Morgan et al., 1989; Zola-Morgan et al., 1994). However, in examining a longer trajectory, some individuals with MTL damage can exhibit preserved, albeit much slower, learning (Bayley et al., 2005; Buffalo et al., 1998; Chudasama et al., 2008; Gaffan & Murray, 1992; Hood et al., 1999; Malamut et al., 1984; Phillips et al., 1988; Suzuki et al., 1993). However, neuroimaging studies of longer learning trajectories in healthy individuals have fallen short of establishing a clear role for the MTL in decision-making (Tremel et al., 2016, 2018). The present study suggests that these findings can be reconciled with the view that MTL is essential for rapidly learning about an initial experience and supporting initial decision-making. This is why model-based learners were most successful on measures of decision-making and learning relative to model-free learners.

If learning were to continue, predictions from a model-free system would become more reliable, supplanting model-based predictions from declarative memory, which require more effort to retrieve (Eichenbaum, 2001; Jacoby, 1991; Mandler, 1980). In other words, while the mnemonic scaffold built by the MTL helps to produce initial decision successes, it becomes less essential as model-free predictions are made more veridical. The goal for repeated decisions is less about recalling specific details about individual prior instances and more about producing a particular response. Thus, model-free reinforcement learning via the striatum can use the initial

decision successes supported by the MTL scaffold to reinforce and automatize correct responses through repetition, eventually producing a habitual action to a rewarding stimulus.

This transition between systems may happen relatively quickly, wherein humans only require about four repetitions to reach 80% accuracy with a large set of items (Tremel et al., 2016, 2018). As such, studies of extended learning may miss these early MTL contributions because they are relatively transient compared to longer model-free learning. Importantly, contextual and structural factors of a task and its goals may influence this transition. This may explain why learning is preserved in some cases of MTL damage, wherein patients may be able to rely on model-free processes and slowly learn about decision-relevant information.

3.3.3 Model-free and model-based systems may interact and integrate to support decision-making

Some evidence suggests that while these systems may normally operate in parallel, they may also interact in particular situations. Neural signatures of model-based learning can overlap with those of model-free learning (Doll et al., 2012), suggesting that these processes are not only difficult to separate but also may influence one another. In fact, hybrid models combining the two approaches have been found to correlate with activity in sub-regions of the striatum and the prefrontal cortex (Gläscher et al., 2010). This suggests that both techniques may be approximations of the actual underlying mechanisms in particular regions, reflecting the possibility that these computations are integrated at some level. Thus, future work to determine the extent to which the brain actually implements reinforcement learning mechanisms is warranted.

In some instances, interactions between MTL and striatal systems may be cooperative, wherein hippocampal model-based replay can enhance model-free learning (Johnson & Redish, 2005) or striatal dopamine can enhance model-based learning ability (Sharp et al., 2016). Connectivity between the hippocampus and striatum can facilitate the utilization of state and reward experience during a decision, suggesting that decisions may be best optimized by a dual-system approach (Wimmer & Shohamy, 2012). However, interactions may also be competitive, as continuous engagement in model-based learning may protect against the formation of habits, via a model-free approach (Gillan et al., 2015). This again suggests that both model-based and model-free processes are subject to influence from contextual and structural factors of a task or goal. This is consistent with the findings in the present study indicating that the MTL and striatum are modulated by the experimental factors that were introduced (reinforcement magnitude and associative context).

As such, interactions between model-based and model-free processes seem to be indirect. Model-based processes via the MTL may be critical to ensure successful outcomes after just one or a few decision experiences. These model-based approaches may allow an individual to continue to experience positive decision outcomes that can help refine model-free predictions about which choice options are best. This would allow a model-free system to build reliable predictions about decision outcomes based on a history of experience. Across extended learning, these predictions would become more veridical and more capable of producing rapid and accurate habitual responses (Seger & Spiering, 2011; Tremel et al., 2016). As this continues, effortful model-based approaches may give way to less effortful model-free approaches as long as successful outcomes can be ensured.

3.3.4 Conclusions

In this study, I hypothesized that a model-based learning system in the MTL is essential to making decisions after one learning experience. I found that individuals who preferred model-based approaches over model-free ones were most successful on measures of decision-making and subsequent episodic memory. Additionally, model-based reinforcement learning was associated with neural substrates in the MTL, while model-free reinforcement learning was supported by the striatum. Model-based learners exhibited generally more activation of MTL regions versus striatal ones and vice versa for model-free learners. These findings indicated that the MTL may establish an initial mnemonic scaffold that supports decision-making after a single experience. Altogether, this study provides additional support for the notion that multiple memory systems underlie successful decision-making and that the MTL in particular may be critical in supporting rapid learning from single exposures.

4.0 GENERAL DISCUSSION

When making a decision, multiple learning and memory systems are drawn upon to help guide behavior based on prior experience. The goal of this dissertation was to investigate the extent to which learning and memory systems supported by the MTL and striatum are co-active early in the learning trajectory of single-exposure deterministic decision-making. Findings indicted that the MTL and striatum contribute to decision-making in parallel and that the MTL is especially important to supporting successful decision-making with a single exposure of experience. The empirical work presented here is a part of larger research program which has highlighted the role of the striatum in supporting extended deterministic learning after many repetitions. Altogether, this research program has highlighted that systems centered in the MTL and the striatum are co-active across a trajectory of learning and together support successful decision-making behavior.

4.1 MTL AND STRIATAL SYSTEMS ARE DIFFERENTIALLY ENGAGED ACROSS THE TRAJECTORY OF LEARNING

The MTL and striatum seem to be co-active in deterministic learning, but each system seems to support different cognitive functions that operate at different stages of learning. On one hand, the MTL system supports the retrieval of instances of prior experience. This system implements a model-based approach to learning wherein memory of prior experience can be used to predict the

outcome of subsequent choices. The MTL seems to be especially effective and preferentially engaged during the initial stages of learning, when an individual has had few opportunities to learn about a decision. On the other hand, a striatal system builds generalized predictions about decision outcomes representing a past history of accumulated experience. This system engages in a model-free learning approach that excels at automatizing and generalizing responses across many repetitions of a decision. Together, these systems work in parallel to optimize decision-making behavior across learning.

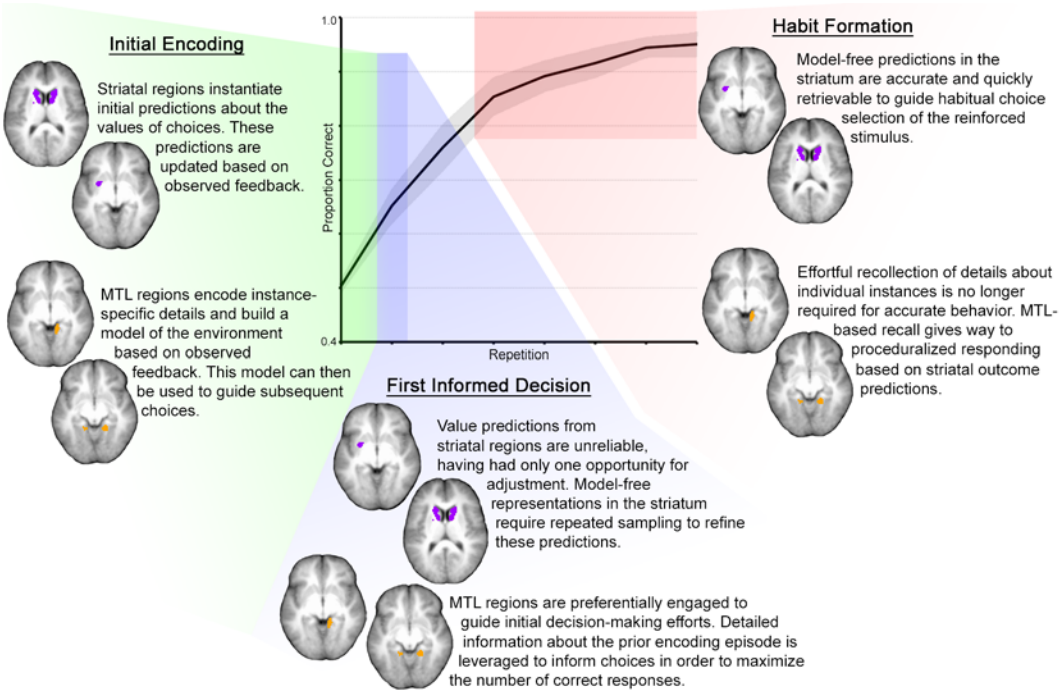

Figure 15. Optimization of decision-making via multiple memory systems.

This schematic illustrates different stages of an extended learning process. The learning curve represents a typically trajectory of improvement across several repetitions of a deterministic decision-making task. During the first exposure to this task (green, Round 1), MTL and striatal regions lay a foundation for further improvement. During the subsequent round (blue), individuals rely on detailed information from the prior encoding experience retrieved via an MTL system. As learning continues (red), predictions from the striatum become more accurate, and decision-making shifts from the initial MTL focus to a striatal focus.

During initial learning stages (e.g., after a single experience), the MTL encodes information about a decision episode and its associated outcome to lay a mnemonic foundation that can be drawn upon to make a successful subsequent decision. In the present study, this functional role was observed in that the MTL supported the encoding of episodic information about individual items and predicted subsequent decision accuracy. While activity in the striatum was also found to support subsequent decision behavior, individuals who were most successful in this initial learning stage relied mostly on MTL-based processes (i.e., model-based reinforcement learning). These findings align with the idea that striatal reinforcement learning processes (i.e., model-free) require repeated sampling of decision outcomes before this system is capable of providing accurate predictions to drive behavior (Daw & Doya, 2006; Dayan & Daw, 2008; O'Doherty et al., 2015; Seger & Spiering, 2011; Seger et al., 2010; Tremel et al., 2016; Tricomi et al., 2009; Yin et al., 2004, 2005). In this sense, the MTL is necessary to scaffold this early behavior and ensure initial decision successes.

This notion of an MTL-based scaffold is important for the striatal system. The striatum is especially sensitive to positive reward prediction errors and is most effective at supporting learning associated with positive outcomes (e.g., monetary reward). Therefore, in order for the striatum to best support learning, an individual would need to make as many correct responses as possible. The MTL can ensure these initial successes via instance-based retrieval so that the striatal system can build accurate value predictions and automatize rewarding decision responses. As such, the striatal system can prepare for a longer learning trajectory by updating initial value predictions with relevant early experiences. Thus, while the MTL and striatal systems may not interact directly, the behavior supported by this MTL scaffold can impact the later success of the striatal system.

As learning continues and decisions are repeated, predictions from the striatum become more reliable. Prior work in this research program has illustrated that regions in the dorsal striatum support successful learning during repetitive decision-making (Tremel et al., 2016, 2018). These studies have also failed to associate MTL regions with this longer learning trajectory, suggesting that while it is essential to initial success, the burden of processing may shift from the MTL to a striatal system. As such, instance-based retrieval via the MTL at this later stage may involve the retrieval of more information than is needed to correctly respond to a particular stimulus. For repeated decisions, the goal is to enact the correct behavior rather than reconstruct the details about each particular prior experience of that decision. As such, this shift from MTL- to striatal-based processing helps to proceduralize these responses so that they can become habitual and less effortful (Seger & Spiering, 2011; Seger et al., 2010; Tremel et al., 2016; Tricomi et al., 2009).

However, while the striatal system seems to eventually overtake the initial contributions of the MTL in a longer learning trajectory, the consequences of learning may still depend on the extent to which each system was engaged across learning. When the efficacy of MTL-based learning is reduced, the resulting memory after a period of repeated learning seems to resemble stimulus-outcome associations produced by the striatum (Tremel et al., 2018). This suggests that if there is a greater reliance on the striatal system during learning, decisions may become more automatic or habitual, but this may come at the cost of a decreased ability to recall instance-specific information about details of prior experiences. Likewise, continual engagement of the MTL system may protect against the formation of habitual responses, but this typically requires more time and effort to retrieve enough detail to guide behavior (Gillan et al., 2015). Taken together, this suggests that the relative engagement of the MTL and striatum across a trajectory

of learning can impact the kinds of representations that are retained later. As such, there is likely an optimal balance between automaticity and ability to retrieve details, but that may depend on the goals of a particular task, process, or individual.

4.2 CONTEXTUAL FACTORS CAN INFLUENCE HOW INDIVIDUALS LEARN AND WHICH SYSTEMS ARE ENGAGED

A common theme that has emerged from this research program is that the relative engagement of the MTL and striatum is highly context dependent. Some factors can increase the relative recruitment of the MTL system, such as visual imagery (Hannula et al., 2013; Hayes et al., 2010; Howard et al., 2011; Park et al., 2014). Other factors can increase the relative recruitment of the striatum, such as list-length, positive reinforcement, or subjective reward magnitude (Jocham et al., 2011; Packard & Knowlton, 2002; Schönberg et al., 2007; Wimmer et al., 2012). Different types of factors can influence how these systems are recruited during decision-making and learning.

In the present study, there was a functional anti-correlation between the MTL (i.e., a region in the hippocampus) and the striatum (i.e., a region in the caudate head), suggesting that these system are engaged differentially. This anti-correlation depended on two factors. First, the MTL was more engaged than the striatum for trials with clear landscape images, but there was no distinction for trials with noise-degraded images. Second, the MTL was more engaged than the striatum for trials with low reinforcement magnitude, but there was no distinction for trials with high reinforcement magnitude. These findings are in line with the predictions that these contextual factors should influence the relative engagement of these systems. In particular, the

clear imagery was intended to boost hippocampal engagement, while high reward magnitudes were intended to boost striatal engagement (though, instead the hippocampus was more active for the low reward magnitude trials). Thus, while these systems may be differentially engaged on some trials depending on contextual factors, they seem to be co-active on others. This aligns with the idea that the MTL builds a mnemonic scaffold to support initial learning in that for trials wherein the striatum is less engaged, the MTL seems to exhibit greater engagement. On other trials, they are co-active, suggesting that the MTL is always engaged at this stage of learning to some extent, and that the striatum may provide supplementary information when contextual factors either dampen MTL influence (i.e., noise-degraded images) or enhance striatal influence (i.e., high reward magnitude).

This view that system-level engagement can be modulated by contextual or structural factors may help reconcile apparent discrepancies in the memory and decision-making literatures. Typically, MTL damage will abolish the ability to perform deterministic learning tasks, such as concurrent discrimination learning (Buffalo et al., 1999; Corkin, 2002; Hood et al., 1999; Squire et al., 1988; Squire and Zola, 1996; Zola-Morgan et al., 1989; Zola-Morgan et al., 1994). However, some learning can still be observed in some cases of MTL damage (Bayley et al., 2005; Buffalo et al., 1998; Chudasama et al., 2008; Gaffan & Murray, 1992; Hood et al., 1999; Malamut et al., 1984; Phillips et al., 1988; Suzuki et al., 1993). It may be the case that in the absence of the MTL, the striatum can support learning, but must acquire many repetitions of experience before outcome predictions are accurate enough to guide behavior. Without the retrieval of information from individual instances, individuals must rely on building a history of experience via model-free reinforcement learning. This may be accomplished by systems like the striatum (Bayley et al., 2005; Poldrack et al., 1999; Seger & Spiering, 2011; Seger et al., 2010;)

or other systems involved in the acquisition of semantic information (Duff et al., 2006; Kan et al., 2009; O'Kane et al., 2004; Sharon et al., 2011), which each may utilize a model-free learning process and may involve overlapping neural systems (Ashby & Maddox, 2005; Davies et al., 2009; Poldrack et al., 1999; Scimeca & Badre, 2012). When structural or contextual elements of a task are altered to discourage the use of MTL-based approaches or boost striatal influence (e.g., long retention intervals or large stimulus sets), individuals with MTL damage seem to reach criterion levels of performance relative to controls at a similar rate, though initial learning is much worse (Bayley et al., 2005; Buffalo et al., 1998; Chudasama et al., 2008; Malamut et al., 1984; Phillips et al., 1988; Suzuki et al., 1993; Gaffan & Murray, 1992; Teng et al., 2000; Turchi et al., 2010). Thus, by using factors that discourage instance retrieval or encourage habit learning, one may shift the burden of processing from an MTL-focused approach to a striatal one.

The idea that different factors can influence how these systems respond during a task highlights a possibility that choice experiences could be engineered to maximize the contributions of one system or another. This seems to be the case in cases of MTL damage wherein altering context or structure can putatively boost the recruitment of alternative systems. However, there may be practical implications, wherein this could be used to benefit individuals with disorders or diseases that affect either the MTL system (e.g., mild cognitive impairment or Alzheimer's disease) or striatal system (e.g., Parkinson's or Huntington's diseases). In these situations, it may be possible to alter the factors surrounding patients' everyday experiences to leverage the use of the unaffected system in circumstance in which the afflicted system would typically excel. For example, emphasizing awareness during procedural learning may help Parkinson's patients acquire habit-like responses that would typically be acquired via striatal

reinforcement learning (Moody et al., 2010). While there has been progress in identifying and understanding the factors that influence these different learning systems, it is yet unclear how these factors can be leveraged and applied to these more practical situations.

4.3 CONCLUSIONS

Altogether, this research program has provided evidence that multiple learning and memory systems, centered in the MTL and striatum, are engaged in parallel across a trajectory of learning. The MTL system excels at encoding and retrieving individual episodes of experience to guide decision behavior. This is especially important to support initial decision success, wherein an individual has little experience or history with a particular stimulus and outcome. In contrast, the striatal system excels at refining decision behavior with repetition to increase response automaticity and form habits. This is important in decreasing the effort and time associated with responding to a particular decision and to regularize repeated experiences. Both of these systems together underlie successful deterministic decision-making and dynamically operate across different stages of learning.

APPENDIX A

FULL STATISTICS FOR MIXED LOGIT ACCURACY MODELS

Table 11. Statistics for mixed logit accuracy models.

ROI	*Effect*	*Coef.* β	*SE* β	*z*	*P*	p_{FDR}	
R Hippocampus	Intercept	1.19	0.28	4.27	<0.001	<0.001	*
	Activity	-0.29	0.13	-2.23	0.03	<0.05	*
	R1 Outcome	-0.89	0.34	-2.67	<0.01	0.02	*
	Activity*Outcome	0.12	0.15	-0.78	0.43	0.61	
Bilat. Parahippocampal G.	Intercept	1.20	0.28	4.36	<0.001	<0.001	*
	Activity	0.02	0.25	0.06	0.95	0.97	
	R1 Outcome	-0.88	0.32	-2.74	<0.01	0.02	*
	Activity*Outcome	-0.41	0.37	-1.13	0.26	0.38	
L Putamen	Intercept	1.18	0.26	4.49	<0.001	<0.001	*
	Activity	0.25	0.15	1.71	0.09	0.14	
	R1 Outcome	-0.87	0.33	-2.66	<0.01	0.02	*
	Activity*Outcome	-0.49	0.21	-2.38	0.02	0.03	*
Bilat. Head of caudate	Intercept	1.18	0.26	4.60	<0.001	<0.001	*
	Activity	0.05	0.32	0.16	0.87	0.97	
	R1 Outcome	-0.85	0.32	-2.67	<0.01	0.02	*
	Activity*Outcome	-0.20	0.42	-0.49	0.62	0.80	
L Hippocampus	Intercept	1.16	0.26	4.45	<0.001	<0.001	*
	Activity	-0.04	0.15	-0.23	0.82	0.96	
	R1 Outcome	-0.83	0.31	-2.67	<0.01	0.02	*
	Activity*Outcome	0.09	0.21	0.42	0.67	0.82	
L Body of caudate	Intercept	1.19	0.26	4.59	<0.001	<0.001	*
	Activity	-0.01	0.16	-0.03	0.97	0.97	
	R1 Outcome	-0.85	0.31	-2.71	<0.01	0.02	*
	Activity*Outcome	-0.01	0.21	-0.06	0.95	0.97	
L Parahippocampal G.	Intercept	1.16	0.26	4.52	<0.001	<0.001	*
	Activity	-0.05	0.10	-0.48	0.63	0.80	
	R1 Outcome	-0.88	0.32	-2.72	<0.01	0.02	*
	Activity*Outcome	0.15	0.14	1.12	0.26	0.38	

APPENDIX B

FULL STATISTICS FOR MIXED LINEAR RT MODELS

Table 12. Statistics for mixed linear response time models.

ROI	*Effect*	*Coef.* β	*SE* β	*z*	*p*	p_{FDR}	
R Hippocampus	Intercept	1.98	0.07	27.89	<0.001	<0.001	*
	Activity	-0.02	0.02	-1.04	0.30	0.46	
	R1 Outcome	0.07	0.03	2.22	0.03	0.07	
	Activity*Outcome	0.01	0.03	0.35	0.73	0.76	
Bilat. Parahippocampal G.	Intercept	1.98	0.07	28.00	<0.001	<0.001	*
	Activity	0.14	0.06	2.47	0.01	0.04	*
	R1 Outcome	0.06	0.03	2.12	0.03	0.07	
	Activity*Outcome	0.06	0.08	0.75	0.45	0.58	
L Putamen	Intercept	1.99	0.07	27.82	<0.001	<0.001	*
	Activity	-0.03	0.03	-0.83	0.41	0.54	
	R1 Outcome	0.06	0.03	2.09	0.04	0.07	
	Activity*Outcome	0.03	0.05	0.68	0.50	0.61	
Bilat. Head of caudate	Intercept	1.98	0.07	28.25	<0.001	<0.001	*
	Activity	0.21	0.07	3.02	<0.01	<0.01	*
	R1 Outcome	0.06	0.03	2.06	0.04	0.07	*
	Activity*Outcome	0.05	0.09	0.57	0.57	0.65	
L Hippocampus	Intercept	1.98	0.07	27.83	<0.001	<0.001	*
	Activity	0.03	0.03	0.94	0.35	0.51	
	R1 Outcome	0.06	0.03	2.15	0.03	0.07	
	Activity*Outcome	-0.02	0.05	-0.51	0.61	0.66	
L Body of caudate	Intercept	1.98	0.07	28.73	<0.001	<0.001	*
	Activity	-0.03	0.04	-0.91	0.36	0.51	
	R1 Outcome	0.06	0.03	1.91	0.06	0.10	
	Activity*Outcome	0.09	0.05	1.85	0.06	0.11	
L Parahippocampal G.	Intercept	1.99	0.07	27.68	<0.001	<0.001	*
	Activity	0.00	0.02	0.18	0.86	0.86	
	R1 Outcome	0.07	0.03	2.22	0.03	0.07	
	Activity*Outcome	-0.02	0.03	-0.56	0.58	0.65	

APPENDIX C

FULL ANOVA TABLE FOR RECOGNITION ITEM ANALYSIS

Table 13. Statistics for recognition memory item analysis ANOVAs.

ROI	*Effect*	*F*	*p*	p_{FDR}	
R Hippocampus	Memory Accuracy	2.46	0.12	0.55	
	Memory Accuracy*R1 Outcome	0.33	0.57	0.65	
	Memory Accuracy*Reward	8.93	0.003	<0.05	*
	Memory Accuracy*Context	0.71	0.40	0.59	
	Memory Accuracy*R1 Outcome*Reward	1.08	0.30	0.56	
	Memory Accuracy*R1 Outcome*Context	1.61	0.20	0.55	
	Memory Accuracy*Reward*Context	0.86	0.35	0.59	
	Memory Accuracy*R1 Outcome*Reward*Context	1.78	0.18	0.55	
Bilat. Parahippocampal G.	Memory Accuracy	3.54	0.06	0.49	
	Memory Accuracy*R1 Outcome	1.92	0.17	0.55	
	Memory Accuracy*Reward	0.74	0.39	0.59	
	Memory Accuracy*Context	3.50	0.06	0.49	
	Memory Accuracy*R1 Outcome*Reward	2.31	0.13	0.55	
	Memory Accuracy*R1 Outcome*Context	0.80	0.37	0.59	
	Memory Accuracy*Reward*Context	0.69	0.41	0.59	
	Memory Accuracy*R1 Outcome*Reward*Context	0.54	0.46	0.59	
L Putamen	Memory Accuracy	0.30	0.59	0.65	
	Memory Accuracy*R1 Outcome	0.57	0.45	0.59	
	Memory Accuracy*Reward	1.47	0.23	0.55	
	Memory Accuracy*Context	1.37	0.24	0.55	
	Memory Accuracy*R1 Outcome*Reward	0.37	0.55	0.65	
	Memory Accuracy*R1 Outcome*Context	1.47	0.23	0.55	
	Memory Accuracy*Reward*Context	1.12	0.29	0.56	
	Memory Accuracy*R1 Outcome*Reward*Context	0.54	0.46	0.59	
Bilat. Head of caudate	Memory Accuracy	9.27	0.002	<0.05	*
	Memory Accuracy*R1 Outcome	1.13	0.29	0.56	
	Memory Accuracy*Reward	1.79	0.18	0.55	
	Memory Accuracy*Context	0.18	0.67	0.69	
	Memory Accuracy*R1 Outcome*Reward	0.40	0.52	0.65	
	Memory Accuracy*R1 Outcome*Context	0.03	0.85	0.85	
	Memory Accuracy*Reward*Context	2.78	0.10	0.55	
	Memory Accuracy*R1 Outcome*Reward*Context	0.19	0.66	0.69	

APPENDIX D

FULL ANOVA TABLE FOR EPISODIC ITEM ANALYSIS

Table 14. Statistics for episodic memory item analysis ANOVAs.

ROI	*Effect*	*F*	*p*	p_{FDR}	
R Hippocampus	Memory Response	0.00	1.00	0.99	
	Memory Response*R1 Outcome	13.84	<0.001	0.007	**
	Memory Response*Reward	0.21	0.65	0.87	
	Memory Response*Context	0.07	0.80	0.99	
	Memory Response*R1 Outcome*Reward	0.43	0.51	0.87	
	Memory Response*R1 Outcome*Context	2.31	0.13	0.57	
	Memory Response*Reward*Context	0.20	0.65	0.87	
	Memory Response*R1 Outcome*Reward*Context	0.84	0.36	0.87	
Bilat. Parahippocampal G.	Memory Response	0.26	0.61	0.87	
	Memory Response*R1 Outcome	0.21	0.64	0.87	
	Memory Response*Reward	1.42	0.23	0.68	
	Memory Response*Context	8.38	<0.01	<0.05	*
	Memory Response*R1 Outcome*Reward	0.50	0.48	0.87	
	Memory Response*R1 Outcome*Context	8.97	<0.01	<0.05	*
	Memory Response*Reward*Context	2.24	0.13	0.57	
	Memory Response*R1 Outcome*Reward*Context	0.49	0.48	0.87	
L Putamen	Memory Response	3.84	0.05	0.40	
	Memory Response*R1 Outcome	0.78	0.38	0.87	
	Memory Response*Reward	0.50	0.48	0.87	
	Memory Response*Context	2.16	0.14	0.57	
	Memory Response*R1 Outcome*Reward	1.06	0.30	0.81	
	Memory Response*R1 Outcome*Context	0.02	0.89	0.99	
	Memory Response*Reward*Context	0.01	0.90	0.99	
	Memory Response*R1 Outcome*Reward*Context	0.00	0.99	0.99	
Bilat. Head of caudate	Memory Response	3.45	0.06	0.41	
	Memory Response*R1 Outcome	1.68	0.20	0.68	
	Memory Response*Reward	0.05	0.83	0.99	
	Memory Response*Context	0.01	0.95	0.99	
	Memory Response*R1 Outcome*Reward	0.01	0.94	0.99	
	Memory Response*R1 Outcome*Context	1.54	0.21	0.68	
	Memory Response*Reward*Context	0.41	0.52	0.87	
	Memory Response*R1 Outcome*Reward*Context	0.24	0.62	0.87	

APPENDIX E

EXAMPLE OF UPDATING FOR THE MODEL-BASED TRANSITION MATRIX

To illustrate how the transition probability matrix of the model-based reinforcement learning agent is learned, this example will walk through two trials from a single subject. One trial received a monetary gain during Round 1, while the other received a monetary loss. For this example, the learning rate, η, is set to 0.80. In Round 1, transition probabilities for all trials are instantiated at 0.5, since subjects have no information about the real probabilities. Each possible Round 1 action, A, has a transition matrix:

Table 15. Initial transition probability matrix.

S_{R1}	S_{R2}	S'_{R2}
1	0.5	0.5
2	0.5	0.5

where S_{R1} is the Round 1 state (i.e., trial number), and S_{R2} and S'_{R2} are the two possible states that can be experienced in Round 2. For the first trial, $S_{R1} = 1$, a monetary gain is received, and a state prediction error is computed using Equation 3:

$$\delta_{SPE}(gain) = 1 - T(s, a, s')$$

$$\delta_{SPE}(gain) = 1 - 0.5 = 0.5$$

This state prediction error is then used to update the transition probability using Equation 5:

$$T(s, a, s') = T(s, a, s') + \eta \times \delta_{SPE}$$

$$T(s, a, s') = 0.5 + 0.8 \times 0.5 = 0.9$$

The unvisited state is also updated via Equation 6:

$$T(s, a, s') = T(s, a, s') \times (1 - \eta)$$

$$T(s, a, s') = 0.5 \times (1 - 0.8) = 0.1$$

Thus, the updated transition matrix after this first trial looks like the following:

Table 16. First trial update to the transition matrix.

S_{R1}	S_{R2}	S'_{R2}
1	0.9	0.1
2	0.5	0.5

This process is repeated for the second trial, though since this trial is associated with monetary loss, different equations are used. First, the state prediction error is computed using Equation 4:

$$\delta_{SPE}(loss) = 0 - T(s, a, s')$$

$$\delta_{SPE}(loss) = 0 - 0.5 = -0.5$$

which is then used to update the initial transition probability via Equation 5:

$$T(s, a, s') = T(s, a, s') + \eta \times \delta_{SPE}$$

$$T(s, a, s') = 0.5 + 0.8 \times (-0.5) = 0.1$$

The transition probability of the unvisited state is also updated, using Equation 7:

$$T(s, a, s') = T(s, a, s') \times (1 + \eta)$$

$$T(s, a, s') = 0.5 \times (1 + 0.8) = 0.9$$

This process produces the following transition matrix after this first series of updates:

Table 17. Second trial update to the transition matrix.

S_{R1}	S_{R2}	S'_{R2}
1	0.9	0.1
2	0.1	0.9

After learning this transition matrix for each of the trials, the agent then computes state-action values, *Q*, based on the actual reward received and the associated transition of the state-action pair. These state-action values are used to guide the behavior of the agent.

BIBLIOGRAPHY

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-178.

Austin, P. C., & Steyerberg, E. W. (2012). Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. *BMC medical research methodology, 12*(1), 82.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445-459.

Bartoń, L. (2017). MuMIn: Multi-Model Inference. R package version 1.40.0. https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bayley, P. J., Frascino, J. C., & Squire, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature, 436*(7050), 550.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289–300.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165–1188.

Bornstein, A. M., & Daw, N. D. (2011). Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Current opinion in neurobiology, 21*(3), 374-380.

Bornstein, A. M., & Daw, N.D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, 35(7), 1011-1023.

Bornstein, A. M., & Daw, N. D. (2013). Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS computational biology, 9*(12), e1003387.

Buffalo, E. A., Stefanacci, L., Squire, L. R., & Zola, S. M. (1998). A reexamination of the concurrent discrimination learning task: the importance of anterior inferotemporal cortex, area TE. *Behavioral neuroscience, 112*(1), 3.

Buffalo, E. A., Ramus, S. J., Clark, R. E., Teng, E., Squire, L. R., & Zola, S. M. (1999). Dissociation between the effects of damage to perirhinal cortex and area TE. *Learning & Memory, 6*(6), 572-599.

Burnham, K. P., & Anderson, D. R. (2003). Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

Camille, N., Tsuchida, A., & Fellows, L. K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *Journal of Neuroscience, 31*(42), 15048-15052.

Chalmers, E., Luczak, A., & Gruber, A. J. (2016). Computational properties of the hippocampus increase the efficiency of goal-directed foraging through hierarchical reinforcement learning. *Frontiers in computational neuroscience, 10*, 128.

Chudasama, Y., Wright, K. S., & Murray, E. A. (2008). Hippocampal lesions in rhesus monkeys disrupt emotional responses but not reinforce devaluation effects. *Biological psychiatry, 63*(11), 1084-1091.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*(2), 291-307.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125*(5), 1054-1069.

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience, 27*(2), 371-378.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505.

Corkin, S. (2002). What's new with the amnesic patient HM? *Nature Reviews Neuroscience, 3*(2), 153.

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of neurophysiology, 88*(2), 982-990.

Davies, R. R., Halliday, G. M., Xuereb, J. H., Kril, J. J., & Hodges, J. R. (2009). The neural basis of semantic memory: Evidence from semantic dementia. *Neurobiology of aging, 30*(12), 2043-2052.

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current opinion in neurobiology, 16*(2), 199-204.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204-1215.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience, 8*(4), 429-453.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of neurophysiology, 84*(6), 3072-3077.

Delgado, M. R., & Dickerson, K. C. (2012). Reward-related learning via multiple memory systems. *Biological psychiatry, 72*(2), 134-141.

Dickerson, K. C., Li, J., & Delgado, M. R. (2011). Parallel contributions of distinct human memory systems during probabilistic learning. *Neuroimage, 55*(1), 266-276.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology, 22*(6), 1075-1081.

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature neuroscience, 18*(5), 767.

Doll, B. B., Shohamy, D., & Daw, N. D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiology of learning and memory, 117*, 4-13.

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature neuroscience, 9*(1), 140.

Duncan, K., Doll, B. B., Daw, N. D., & Shohamy, D. (2018). More Than the Sum of Its Parts: A Role for the Hippocampus in Configural Reinforcement Learning. *Neuron, 98*(3), 645-657.

Eichenbaum, H. (2001). The hippocampus and declarative memory: cognitive mechanisms and neural codes. *Behavioural brain research, 127*(1-2), 199-207.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of neuroscience, 20*(16), 6159-6165.

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron, 76*(6), 1057-1070.

FitzGerald, T. H., Seymour, B., & Dolan, R. J. (2009). The role of human orbitofrontal cortex in value comparison for incommensurable objects. *Journal of Neuroscience, 29*(26), 8388-8395.

Fox, M. D., Snyder, A. Z., Barch, D. M., Gusnard, D. A., Raichle, M. E. (2005). Transient BOLD responses at block transitions, *Neuroimage, 28*, 956-966.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*. *Sage Press*. Los Angeles: Thousand Oaks.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual review of psychology, 49*(1), 87-115.

Gaffan, D., & Murray, E. A. (1992). Monkeys (Macaca fascicularis) with rhinal cortex ablations succeed in object discrimination learning despite 24-hr intertrial intervals and fail at matching to sample despite double sample presentations. *Behavioral neuroscience, 106*(1), 30.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual review of psychology, 68*, 101-128.

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience, 15*(3), 523-536.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585-595.

Hannula, D. E., Libby, L. A., Yonelinas, A. P., & Ranganath, C. (2013). Medial temporal lobe contributions to cued retrieval of items and contexts. *Neuropsychologia, 51*(12), 2322-2332.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of neuroscience, 28*(22), 5623-5630.

Harrell, F. E. (2015). Ordinal logistic regression. *In Regression modeling strategies* (pp. 311-325). Springer, Cham.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., ... & Kawato, M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. Journal of *Neuroscience, 24*(7), 1660-1665.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. Philosophical Transactions of the Royal Society of London B: *Biological Sciences, 364*(1521), 1263-1271.

Hayes, S. M., Baena, E., Truong, T. K., & Cabeza, R. (2010). Neural mechanisms of context effects on face recognition: automatic binding and context shift decrements. *Journal of cognitive neuroscience, 22*(11), 2541-2554.

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature, 431*(7010), 859.

Hood, K. L., Postle, B. R., & Corkin, S. (1999). An evaluation of the concurrent discrimination task as a measure of habit learning: performance of amnesic subjects. *Neuropsychologia, 37*(12), 1375-1386.

Howard, L. R., Kumaran, D., Ólafsdóttir, H. F., & Spiers, H. J. (2011). Double dissociation between hippocampal and parahippocampal responses to object–background context and scene novelty. *Journal of Neuroscience, 31*(14), 5253-5261.

Ito, M., & Doya, K. (2018). Information Coded in the Striatum During Decision-Making. In *Advances in Cognitive Neurodynamics (VI)* (pp. 19-25). Springer, Singapore.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language, 30*(5), 513-541.

Jocham, G., Klein, T. A., & Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *Journal of Neuroscience, 31*(5), 1606-1613.

Johnson, A., & Redish, A. D. (2005). Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Networks, 18*(9), 1163-1171.

Kan, I. P., Alexander, M. P., & Verfaellie, M. (2009). Contribution of prior semantic knowledge to new episodic learning in amnesia. *Journal of Cognitive Neuroscience, 21*(5), 938-944.

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The role of mental maps in decision-making. *Trends in neurosciences, 40*(5), 256-259.

Kim, H. F., & Hikosaka, O. (2013). Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron, 79*(5), 1001-1010.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*(5280), 1399-1402.

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron, 63*(6), 889-901.

Lancaster, J. L., Glass, T. G., Lankipalli, B. R., Downs, H., Mayberg, H., Fox, P. T. (1995). A modality-independent approach to normalization of tomographic images of the human brain. *Human brain mapping, 3*, 209-223.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron, 81*(3), 687-699.

Liu, X., Powell, D. K., Wang, H., Gold, B. T., Corbly, C. R., & Joseph, J. E. (2007). Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. *Journal of Neuroscience, 27*(17), 4587-4597.

Lüdecke, D. (2017). sjPlot: Data Visualization for Statistics in Social Science. R Package Version 2.4.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review, 87*(3), 252.

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature neuroscience, 20*(9), 1269.

Miyachi, S., Hikosaka, O., & Lu, X. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Experimental brain research, 146*(1), 122-126.

Malamut, B. L., Saunders, R. C., & Mishkin, M. (1984). Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behavioral neuroscience, 98*(5), 759.

Mack, M. L., & Preston, A. R. (2016). Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *Neuroimage, 127*, 144-157.

Mahut, H., Zola-Morgan, S., & Moss, M. (1982). Hippocampal resections impair associative learning and recognition memory in the monkey. *Journal of Neuroscience, 2*(9), 1214-1220.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in cognitive sciences, 7*(7), 293-299.

Mishkin, M. (1982). A memory system in the monkey. *Phil. Trans. R. Soc. Lond. B, 298*(1089), 85-95.

Moody, T. D., Chang, G. Y., Vanek, Z. F., & Knowlton, B. J. (2010). Concurrent discrimination learning in Parkinson's disease. *Behavioral neuroscience, 124*(1), 1.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology, 53*(3), 139-154.

O'Doherty, J. P., Lee, S. W., & McNamee, D. (2015). The structure of reinforcement-learning mechanisms in the human brain. *Current Opinion in Behavioral Sciences, 1*, 94-100.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual review of psychology, 68*, 73-100.

O'kane, G., Kensinger, E. A., & Corkin, S. (2004). Evidence for semantic learning in profound amnesia: an investigation with patient HM. *Hippocampus, 14*(4), 417-425.

Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E., & Conturo, T. E. (1997). Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage, 6*(3), 156-167.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual review of neuroscience, 25*(1), 563-593.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441*(7090), 223.

Park, H., Abellanoza, C., Schaeffer, J., & Gandy, K. (2014). Source recognition by stimulus content in the MTL. *Brain research, 1553*, 59-68.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods, 162*(1-2), 8-13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics, 2*, 10.

Pennartz, C. M. A., Ito, R., Verschure, P. F. M. J., Battaglia, F. P., & Robbins, T. W. (2011). The hippocampal–striatal axis in learning, prediction and goal-directed behavior. *Trends in neurosciences, 34*(10), 548-559.

Phillips, R. R., Malamut, B. L., Bachevalier, J., & Mishkin, M. (1988). Dissociation of the effects of inferior temporal and limbic lesions on object discrimination learning with 24-h intertrial intervals. *Behavioural Brain Research, 27*(2), 99-107.

Ploran, E. J., Nelson, S. M., Velanova, K., Donaldson, D. I., Petersen, S. E., & Wheeler, M. E. (2007). Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *Journal of Neuroscience, 27*(44), 11912-11924.

Ploran, E. J., Tremel, J. J., Nelson, S. M., & Wheeler, M. E. (2011). High quality but limited quantity perceptual evidence produces neural accumulation in frontal and parietal cortex. *Cerebral Cortex, 21*(11), 2650-2662.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*(4), 564.

Poldrack, R. A., & Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiology of learning and memory, 82*(3), 324-332.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rehbein, L., Killiany, R., & Mahut, H. (2005). Developmental study of the hippocampal formation in rhesus monkeys (Macaca mulatta): I. Early ablations spare discrimination learning but not recognition memory. *Behavioral neuroscience, 119*(3), 635.

Ridley, R. M., Aitken, D. M., & Baker, H. F. (1989). Learning about rules but not about reward is impaired following lesions of the cholinergic projection to the hippocampus. *Brain research, 502*(2), 306-318.Schacter et al 2012

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27*(47), 12860-12867.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Schultz, W. (2013). Updating dopamine reward signals. *Current opinion in neurobiology, 23*(2), 229-238.

Scimeca, J. M., & Badre, D. (2012). Striatal contributions to declarative memory retrieval. *Neuron, 75*(3), 380-392.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25*(11), 2941-2951.

Seger, C. A., & Spiering, B. J. (2011). A critical review of habit learning and the basal ganglia. *Frontiers in Systems Neuroscience, 5*, 66.

Seger, C. A., Peterson, E. J., Cincotta, C. M., Lopez-Paniagua, D., & Anderson, C. W. (2010). Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. *Neuroimage, 50*(2), 644-656.

Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron, 90*(5), 927-939.

Sharon, T., Moscovitch, M., & Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences, 108*(3), 1146-1151.

Sharp, M. E., Foerde, K., Daw, N. D., & Shohamy, D. (2015). Dopamine selectively remediates 'model-based' reward learning: a computational approach. Brain, 139(2), 355-364.

Snyder, A. Z. (1996). Difference Image vs Ratio Image Error Function Forms in PET—PET Realignment. *In Quantification of brain function using PET* (pp. 131-137).

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review, 99*(2), 195.

Squire, L. R., Zola-Morgan, S., & Chen, K. S. (1988). Human amnesia and animal models of amnesia: performance of amnesic patients on tests designed for the monkey. *Behavioral neuroscience, 102*(2), 210.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, 93*(24), 13515-13522.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*(5026), 1380-1386.

Suzuki, W. A., Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1993). Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *Journal of Neuroscience, 13*(6), 2430-2451.

Suzuki, M., Johnson, J. D., & Rugg, M. D. (2011). Decrements in hippocampal activity with item repetition during continuous recognition: an fMRI study. Journal of Cognitive *Neuroscience, 23*(6), 1522-1532.

Talairach, J., & Tournoux, P. (1988). Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging (Thieme Classics). *Thieme*.

Teng, E., Stefanacci, L., Squire, L. R., & Zola, S. M. (2000). Contrasting effects on discrimination learning after hippocampal lesions and conjoint hippocampal–caudate lesions in monkeys. *Journal of Neuroscience, 20*(10), 3853-3863.

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *Neuroimage, 109*, 35-49.

Tremel, J. J., Laurent, P. A., Wolk, D. A., Wheeler, M. E., & Fiez, J. A. (2016). Neural signatures of experience-based improvements in deterministic decision-making. *Behavioural brain research, 315*, 51-65.

Tremel, J. J., Ortiz, D. M., & Fiez, J. A. (2018). Manipulating memory efficacy affects the behavioral and neural profiles of deterministic learning and decision-making. *Neuropsychologia, 114*, 214-230.

Tricomi, E., & Fiez, J. A. (2008). Feedback signals in the caudate reflect goal achievement on a declarative memory task. *Neuroimage, 41*(3), 1154-1167.

Tricomi, E., Balleine, B.W., O'Doherty, J.P. (2009). A specific role for the posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29*, 2225-2232.

Tricomi, E., & Fiez, J. A. (2012). Information content and reward processing in the human striatum during performance of a declarative memory task. *Cognitive, Affective, & Behavioral Neuroscience, 12*(2), 361-372.

Turchi, J., Devan, B., Yin, P., Sigrist, E., & Mishkin, M. (2010). Pharmacological evidence that both cognitive memory and habit formation contribute to within-session learning of concurrent visual discriminations. *Neuropsychologia, 48*(8), 2245-2250.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review, 108*(3), 550.

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *Journal of Neuroscience, 27*(15), 4019-4026.

Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron, 49*(3), 459-466.

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science, 338*(6104), 270-273.

Wimmer, G. E., Daw, N. D., & Shohamy, D. (2012). Generalization of value in reinforcement learning by humans. *European Journal of Neuroscience, 35*(7), 1092-1104.

Yin, H. H., Knowlton, B. J., Balleine, B. W. (2004). Lesions of the dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience, 19*, 181-189.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22*(2), 513-523.

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 356(1413), 1363-1374.

Yu, S. S., Johnson, J. D., & Rugg, M. D. (2012). Hippocampal activity during recognition memory co-varies with the accuracy and confidence of source memory judgments. *Hippocampus, 22*(6), 1429-1437.

Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1989). Lesions of the hippocampal formation but not lesions of the fornix or the mammillary nuclei produce long-lasting memory impairment in monkeys. *Journal of Neuroscience, 9*(3), 898-913.

Zola-Morgan, S., Squire, L. R., & Ramus, S. J. (1994). Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus, 4*(4), 483-495.