Available online at www.sciencedirect.com

SCIENCE DIRECT®

Neurobiology of Learning and Memory 82 (2004) 253–267

Neurobiology of Learning and Memory

www.elsevier.com/locate/ynlme

Hippocampus, cortex, and basal ganglia: Insights from computational models of complementary learning systems

Hisham E. Atallah, Michael J. Frank, and Randall C. O'Reilly*

Department of Psychology, Center for Neuroscience, University of Colorado at Boulder, 345 UCB, Boulder, CO 80309, USA

Received 16 April 2004; revised 4 June 2004; accepted 8 June 2004
Available online 20 July 2004

Abstract

We present a framework for understanding how the hippocampus, neocortex, and basal ganglia work together to support cognitive and behavioral function in the mammalian brain. This framework is based on computational tradeoffs that arise in neural network models, where achieving one type of learning function requires very different parameters from those necessary to achieve another form of learning. For example, we dissociate the hippocampus from cortex with respect to general levels of activity, learning rate, and level of overlap between activation patterns. Similarly, the frontal cortex and associated basal ganglia system have important neural specializations not required of the posterior cortex system. Taken together, this overall cognitive architecture, which has been implemented in functioning computational models, provides a rich and often subtle means of explaining a wide range of behavioral and cognitive neuroscience data. Here, we summarize recent results in the domains of recognition memory, contextual fear conditioning, effects of basal ganglia lesions on stimulus–response and place learning, and flexible responding.

Keywords: Computational models; Hippocampus; Basal ganglia; Neocortex

1. Introduction

The brain is not a homogenous organ: different brain areas clearly have some degree of specialized function. There have been many attempts to specify what these functions are, based on a variety of theoretical approaches and data. In this paper, we summarize our approach to this problem, which is based on the logic of *computational tradeoffs* in neural network models of brain areas. The core idea behind this approach is that different brain areas are specialized to satisfy fundamental tradeoffs in the way that neural systems perform different kinds of learning and memory tasks. This way of characterizing the specializations of brain areas is in many ways consistent with ideas from other frameworks, but we argue that it offers a level of precision and subtlety that may prove beneficial in understanding complex interactions between different brain areas. This paper reviews a number of illustrations of this point, through applications of computational models to a range of data in both the human and animal literatures, including: recognition memory, contextual fear conditioning, effects of basal ganglia lesions on stimulus–response and place learning, and flexible responding.

One of the central tradeoffs behind our approach involves the process of learning novel information rapidly without interfering catastrophically with prior knowledge. This form of learning requires a neural network with very sparse levels of overall activity (leading to highly separated representations), and a relatively high learning rate. These features are incompatible with the kind of network that is required to acquire general statistical information about the environment, which needs highly overlapping, distributed representations with relatively higher levels of activity, and a slow rate of learning. The conclusion we have drawn from this mutual incompatibility is that the brain must have two different learning systems to perform these different functions, and this fits quite well with a

* Corresponding author. Fax: 1-303-492-2967.
E-mail address: oreilly@psych.colorado.edu (R.C. O'Reilly).

1074-7427/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.nlm.2004.06.004

wide range of converging cognitive neuroscience data on the properties of the hippocampus and posterior neocortex, respectively (e.g., McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly & McClelland, 1994; O'Reilly & Rudy, 2001).

A similar kind of reasoning has been applied to understanding the specialized properties of the frontal cortex (particularly focused on the prefrontal cortex) relative to the posterior neocortex and hippocampal systems. The tradeoff in this case involves specializations required for maintaining information in an active state (i.e., maintained neural firing) relative to those required for performing semantic associations and other forms of inferential reasoning. Specifically, active maintenance (often referred to by the more general term of working memory) requires relatively isolated representations so that information does not spread out and get lost over time (O'Reilly, Braver, & Cohen, 1999; O'Reilly & Munakata, 2000). In contrast, the overlapping distributed representations of posterior cortex support spreading associations and inference by allowing one representation to activate aspects of other related representations. The prefrontal cortex system also requires an adaptive gating mechanism to be able to rapidly update new information while also robustly maintaining other information—the basal ganglia have the right neural properties to provide this function (Frank, Loughry, & O'Reilly, 2001).

Putting these arguments together, this computational framework supports a tripartite cognitive architecture represented in Fig. 1, composed of posterior cortex (PC), hippocampus (HC), and frontal cortex (FC), which is thought to include the basal ganglia as well (and many other relevant brain areas are not included, for simplicity). Each component of the architecture is specialized for a different function by virtue of having different parameters and neural specializations (as motivated by the above tradeoffs), but the fundamental underlying mechanisms are the same across all areas. Specifically, our models are all implemented within the Leabra framework (O'Reilly, 1998; O'Reilly & Munakata, 2000), which includes a coherent set of basic neural processing and learning mechanisms that have been developed by different researchers over the years. Thus, many aspects of these areas work in the same way, and in many respects the system can be considered to function as one big undifferentiated whole. For example, any given memory is encoded in synapses distributed throughout the entire system, and all areas participate in some way in representing most memories. Therefore, this architecture is much less modular than most conceptions of the brain, while still providing a principled and specific way of understanding the differential contributions of different brain areas. These seemingly contradictory statements are resolved through the process of developing and testing concrete computational simulations that help us understand the ways in which these areas contribute differentially, and similarly, to cognitive and behavioral functions.

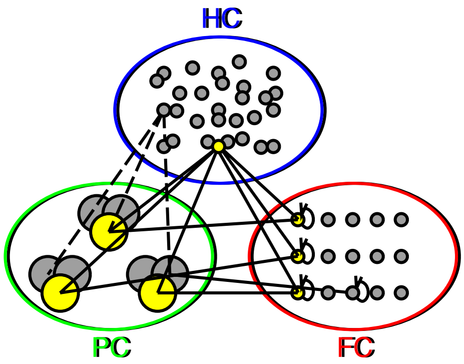


Fig. 1. Tripartite cognitive architecture defined in terms of different computational tradeoffs associated with posterior cortex (PC), hippocampus (HC), and frontal cortex (FC) (with motor frontal cortex constituting a blend between FC and PC specializations). Large overlapping circles in PC represent overlapping distributed representations used to encode semantic and perceptual information. Small separated circles in HC represent sparse, pattern-separated representations used to rapidly encode ("bind") entire patterns of information across cortex while minimizing interference. Isolated, self-connected representations in FC represent isolated stripes (columns) of neurons capable of sustained firing (i.e., active maintenance or working memory). The basal ganglia also play a critical role in the FC system by modulating ("gating") activations there based on learned reinforcement history.

In many ways, the understanding we have achieved through these computational models accords well with theories derived through other motivations. For example, there is broad agreement among theorists that a primary function of the hippocampus is the encoding of episodic or spatial memories (e.g., Squire, 1992; Vargha-Khadem et al., 1997). This function emerges from the use of sparse representations in our models, because these representations cause the system to develop conjunctive representations that bind together the many different features of an episode or location into a unitary encoding (e.g., O'Reilly & McClelland, 1994; O'Reilly & Rudy, 2001). Similarly, a widely held distinction between recognition memory and recall memory in humans (as elaborated later) is supported by our model (Norman & O'Reilly, 2003).

However, the models are also often at variance with existing theorizing. Perhaps the single most pervasive example of this comes from the nearly universal attempts to definitively localize the "engram" or substrate of memory storage. People inevitably want to know, "is this memory in the hippocampus or in the cortex?" As noted above, in our computational models, the answer is always *both* (unless the hippocampus has been removed, of course). Thus, the relevant question is, what kind of behavioral functions can the synaptic changes in one brain area support relative to those in another area? For example, our models show that, with relatively brief exposures, encoding in the hippocampus can often support recall of the specific details of a given episode, while neocortical representations can usually only support a general feeling of familiarity, without the ability to recall specific details (Norman & O'Reilly, 2003). Critically, the traditional notions of "familiarity" and "recall" do not capture all the distinction between neocortical and hippocampal contributions, as we showed in a number of cases in Norman and O'Reilly (2003). For example, neocortical representations can be sensitive to contextual information, and even to arbitrary paired associates, which is not well accounted for by traditional notions of how the familiarity system works.

The distributed nature of memory encoding also bears on the central debates regarding the fate of memories that are initially encoded primarily by the hippocampus—considerable evidence suggests that these memories can become independent of the hippocampus over time through a "consolidation" process (e.g., McClelland et al., 1995; Squire, 1992; Sutherland et al., 2001). Thus, people are tempted to conclude that the memory is "transferred" out of the hippocampus and into the neocortex. In contrast, our models suggest that the neocortical contribution to the memory (which was always present to some degree) is simply strengthened to the point that it becomes capable of more robust recall even in the absence of the hippocampus. This does not mean that the memory has to leave the hippocampus, and indeed we believe that the hippocampal system is actively participating in recalling even very old memories, which is consistent with the theorizing of Moscovitch and Nadel (1998). In short, memory always remains distributed throughout the brain. But different brain areas can support different types of behavioral functions based on their independent encoding of these memories. Labels such as "declarative" and "procedural" do not necessarily capture the subtlety and complexity of these distinctions.

Another example of the subtlety of the computational models comes from understanding the role of the basal ganglia in cognition and behavior. According to our framework, the basal ganglia play an intrinsically *modulatory* role; this can be difficult to accommodate in verbal theories. For example, many people regard the basal ganglia as a "habit learning" system, that learns to encode stimulus–response associations over time (e.g., Packard, Hirsh, & White, 1989). However, it is clear that motor responding is relatively unimpaired by basal ganglia dysfunction; instead, basal ganglia damage seems to affect the ability to initiate or select motor actions (e.g., Hikosaka, 1998; Mink, 1996). Thus, it is more likely that the basal ganglia modulate or gate the functioning of the frontal cortical areas that they project to, helping to select the most appropriate actions for a given situation. The distinction is perhaps a subtle one, but it may have important implications for understanding behavioral data, as we discuss later.

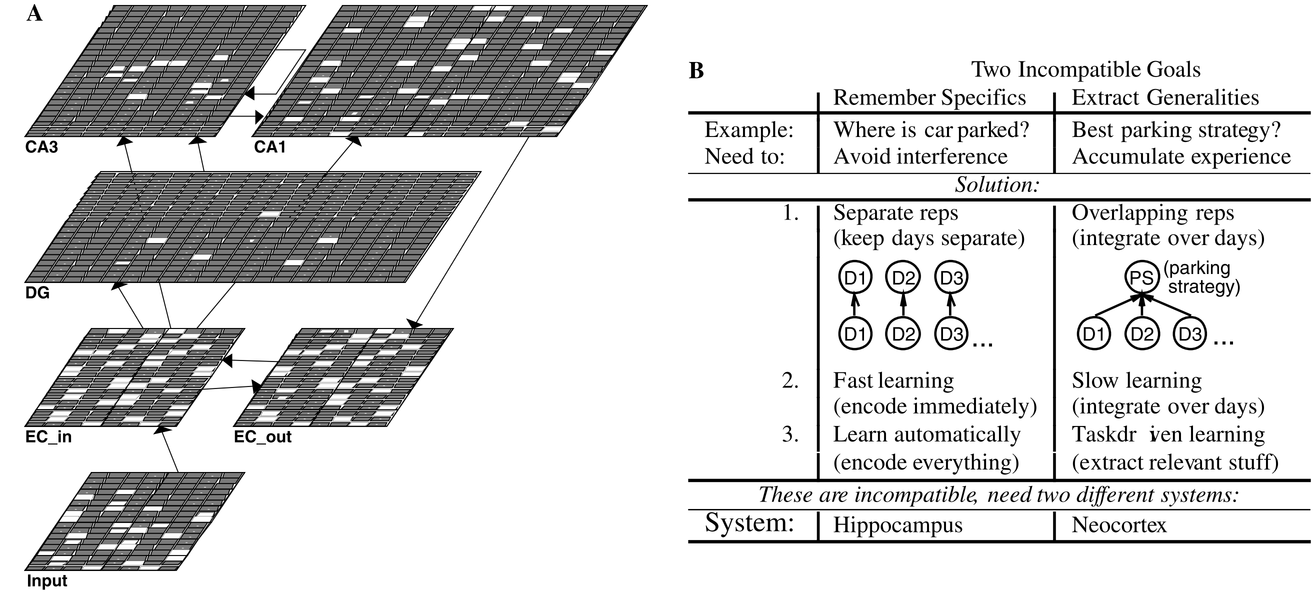
In the remainder of this chapter, we explore the implications of our computational architecture in greater detail, focusing on data regarding the hippocampal contributions to both human and rat learning and memory, and on rat and human studies of basal ganglia function.

2. Hippocampus and posterior neocortex

We have developed an instantiation of our theory in the form of a computational model of the hippocampus and neocortex, as shown in Fig. 2, along with a summary of the computational tradeoff argument. This same basic model has been applied to a wide range of data from animals and humans (Frank, Rudy, & O'Reilly, 2003; Norman & O'Reilly, 2003; O'Reilly, Norman, & McClelland, 1998; O'Reilly & Rudy, 2001; Rudy & O'Reilly, 2001) (see O'Reilly & Norman, 2002 for a concise review). Thus, this model is perhaps one of the most well tested in the neural network literature. As noted earlier, the critical feature of this model is that it employs sparse, though still distributed, representations in the primary hippocampal regions of CA3, CA1, and particularly DG. It also incorporates a number of other features of the hippocampal anatomy and physiology that have been analyzed as being important to its overall functions (O'Reilly & McClelland, 1994).

In brief, the hippocampal model performs encoding and retrieval of memories in the following manner: during encoding, the hippocampus develops relatively non-overlapping (pattern-separated) representations in region CA3 (which is strongly facilitated by the very sparse dentate gyrus (DG) inputs). Active units in CA3 are linked to one another (via Hebbian learning), and to a re-representation of the input pattern in region CA1. During retrieval, presentation of a partial version of a previously encoded memory representation leads to reconstruction of the complete original CA3 representation (i.e., pattern completion) and, through this, reconstruction of the entire studied pattern on the EC output layer (and then to cortex) via area CA1. As reviewed in Norman and O'Reilly (2003) and O'Reilly and Rudy (2001), our hippocampal model closely resembles other neural network models of the hippocampus (Burgess & O'Keefe, 1996; Hasselmo & Wyble, 1997; Moll & Miikkulainen, 1997; Touretzky & Redish, 1996; Treves & Rolls, 1994; Wu, Baxter, & Levy, 1996). There are differences, but the family resemblance between these models far outweighs the differences.

In contrast with the rapid, conjunctive learning supported by the hippocampus, our cortical model can support generalization across a large number of experiences for two main reasons. First, our simulated



	Two Incompatible Goals	
	Remember Specifics	Extract Generalities
Example: Need to:	Where is car parked? Avoid interference	Best parking strategy? Accumulate experience
	Solution:	
1.	Separate reps (keep days separate)	Overlapping reps (integrate over days)
2.	Fast learning (encode immediately)	Slow learning (integrate over days)
3.	Learn automatically (encode everything)	Taskdr iven learning (extract relevant stuff)
	These are incompatible, need two different systems:	
System:	Hippocampus	Neocortex

Fig. 2. (A) Our hippocampus model (Norman & O'Reilly, 2003; O'Reilly & Rudy, 2001), showing an example activity pattern. Note the sparse activity in the DG and CA3, and intermediate sparseness of the CA1—these different levels of sparseness enable rapid learning of arbitrary conjunctive information (i.e., "episodic learning"). (B) Computational motivation for two complementary learning and memory systems in the brain: there are two incompatible goals that such systems need to solve. One goal is to remember specific information (e.g., where one's car is parked). The other is to extract generalities across many experiences (e.g., developing the best parking strategy over a number of different days). The neural solutions to these goals are incompatible: memorizing specifics requires separate representations that are learned quickly, and automatically, while extracting generalities requires overlapping representations and slow learning (to integrate over experiences) and is driven by task-specific constraints. Thus, it makes sense to have two separate neural systems separately optimized for each of these goals.

cortical neurons have a slow learning rate (i.e., small changes in synaptic efficacy after a single presentation of a stimulus). That property insures that any single event has a limited effect on cortical representations. It is the gradual effect of multiple exposures that shapes the representation, which enables these representations to capture things that are reliably present across many experiences (i.e., the general statistical structure or regularities of the environment). Second, our model employs representations that involve a relatively large number of neurons (e.g., roughly 15–25%). This property increases the probability that similar events will activate overlapping groups of neurons. Those shared neurons will be responsible for representing the regularities of the environment across multiple experiences.

2.1. Recognition memory

The hippocampal/cortical model has recently been applied to understand the neural dissociation between recall and recognition (Norman & O'Reilly, 2003). Many converging sources of data support the idea that the cortical areas surrounding the hippocampus (i.e., medial temporal lobe cortex or MTLC, principally the perirhinal cortex) can support a form of recognition memory, but not recall, which depends on the hippocampus proper (for a review, see (Yonelinas, 2002)). In a recognition experiment, a list of stimuli is presented to the subject. At a later time, the subject has to differentiate those stimuli from ones that were not on the list. The hippocampal and cortical neural-network models were presented with inputs (patterns of neural activity) that corresponded to different stimuli. Furthermore, stimulus similarity was manipulated by varying the difference between any two inputs. The effect of these presentations on both the hippocampal and cortical models was measured to determine the unique contributions of the MTLC versus the hippocampus in these kinds of recognition memory tests.

Cortical activation became sharper after repeated presentation of a stimulus. Specifically, the relatively few simulated neurons that were initially the most active increased their activity, while the majority of neurons decreased their activity. This is due to the fact that Hebbian learning increases the synaptic efficacy of highly active units, causing them to be more active, while also suppressing the activation of other neurons through inhibitory interneurons. Thus, it was possible to identify familiar stimuli by measuring the average activity of the units with the highest activity. Familiar stimuli caused a higher average activity in those units than novel stimuli.

Yet the cortical model cannot support recall. Due to its slow learning rate and bias for generalization, the cortex does not encode the details of a stimulus or the context in which it was presented. In fact, when the similarity between a previously seen stimulus and a novel probe stimulus was increased, the cortex failed to reliably differentiate between them. For example, if the studied item was the word "RAT," the cortex produced a high familiarity signal for the plural form "RATS" on the test. As expected from our general theoretical framework, the cortical model generalized across specific instances. If the cortex has a threshold at which a signal is categorized as familiar, both "RAT" and "RATS" may pass that threshold.

It is important to note that by setting a high threshold for recognition, the cortical model may be able to reject the stimulus "RATS" based on its difference from the studied item "RAT." But by doing so, the cortical model risks in rejecting another studied stimulus "CAR" because its familiarity signal did not reach the high recognition threshold. The cortex may incorrectly reject "CAR" on the test because it was not properly encoded in the study phase. In other words, raising the recognition threshold will decrease the number of false alarms (e.g., incorrectly accepting "RATS") but at the same time it will increase the number of misses (e.g., incorrectly rejecting "CAR").

Thus, based on this reasoning, a subject with a focal hippocampal lesion should fail a recognition task if they are asked to either accept or reject stimuli based on their familiarity when the studied and novel stimuli are similar. No matter where the recognition threshold is placed either the false alarms or misses will be impair performance. One such task used in the behavioral literature is the yes/no recognition task. The subject is presented one stimulus on the test, and is asked if the stimulus is familiar or not.

Despite poor performance on the yes/no paradigm, the cortical model performed well on a forced-choice test. On a forced-choice test, the subject is asked to choose between a studied and a very similar novel stimulus (e.g., was "RAT" or "RATS" on the list?). The model produced a reliable difference in familiarity when both the studied and novel item were presented on the test. In this case, there is no need to set a threshold. For every pair of stimuli on the test, the familiarity signals are compared. The stimulus that produces a higher familiarity signal is chosen. Therefore, the model predicts that, contrary to a yes/no paradigm, a forced-choice paradigm will be solvable by subjects with a focal hippocampal lesion.

In contrast with the cortical model, the hippocampal model predicts no difficulties with either the recognition or the recall tests. As we described above, the hippocampus will assign different representations for different stimuli. It will also differentiate similar stimuli based on their specific features. For example, the hippocampus will encode RAT on the study list in its singular form, and will tie the representation to the study context. Accordingly, a subject with an intact hippocampus will be able to recall the stimulus (i.e., when asked to list the studied stimuli). They will also be able to use recall to solve a recognition task (e.g., RAT was on this list not RATS). The ability to recall stimuli will be complemented by the familiarity signal that is independently computed by the MTLC.

2.1.1. Recognition memory after focal hippocampal damage

The above predictions from the computational models have been tested in experiments on a patient with selective hippocampal damage and matched controls. Holdstock et al. (2002) compared recognition performance of patient YR and an age-matched control group. YR is a 61-year-old woman that had focal hippocampal damage due to a painkiller overdose. The damage did not extend to the surrounding MTL cortex. The authors' goal was to determine the conditions under which recognition is spared after focal hippocampal damage. The stimuli were images of different objects. The studied and novel stimuli were in some cases very similar (i.e., images of the same object with minor differences in shape). On the recall test, the subjects had to name the objects they have seen in the study phase. On the yes/no recognition task, images were presented one at a time, and the subjects had to respond "yes" if the image was seen in the study phase. On the forced-choice recognition task, a studied image was presented with two novel ones, and the subjects were asked to find the studied one. The experiment also included a forced-choice object–location association task. In this task, the subjects saw an object placed in a certain location on a table. On the test, the subjects had to recognize the familiar object–location association among a number of novel combinations. All those tasks were matched for difficulty by comparing the control group's performance on each one of them.

YR was impaired on the recall task but not the forced-choice object recognition task. She also showed a deficit compared to the control group on two other recognition tasks. The first task was a yes/no recognition task. But she was impaired on this task only when the studied and novel stimuli were similar. YR was also impaired on a forced-choice object–location association task. As the model predicted, hippocampal damage impaired performance on a yes/no recognition task when the studied and novel stimuli were similar. On the contrary, performance on a forced-choice object recognition task was spared even when the studied and novel stimuli were similar. Thus, the availability of the studied stimulus and the similar novel stimulus on the test increased YR's sensitivity to familiar objects.

Forced-choice performance was impaired only when the task contained stimuli from different modalities. In contrast to the object recognition task, the object–location task involved visual and spatial information.

This finding also supports the model's prediction that the hippocampus is necessary for binding information across modalities.

2.1.2. Contextual conditioning in the rat

Other tests of the computational model have come from experiments conducted on intact and hippocampally lesioned rats. This ability to apply the same model to predict results from human and animal subjects is an important asset, because traditional theories have created a divergence between the human and animal lines of research. For example, concepts like "conscious recollection" and "verbal recall," which are widely used in the human memory literature, but are not applicable to animal research. Some of the most direct tests of the computational models have come from the contextual fear conditioning paradigm, which has been shown to depend on the hippocampus (e.g., Anagnostaras, Maren, & Fanselow, 1999). Contextual fear conditioning refers to the association between an environment and a fear response resulting from an aversive experience in that environment. Typically, a rat is given a mild electric shock in a cage formed of a unique set of cues (shape, size, color, etc.). When at a later time, the rat is returned to this environment, it will show a fear response (freezing). In contrast to the hippocampal involvement in this task, a hippocampal lesion did not impair an association between an auditory cue with shock (e.g., Anagnostaras et al., 1999; Kim & Fanselow, 1992). Thus, the hippocampus seems to be important for the representation of context and not in the learning and expression of a fear response.

The logic of a series of studies designed to test our computational models (Rudy, Barrientos, & OReilly, 2002; Rudy & O'Reilly, 1999, 2001), builds on the pre-exposure version of contextual fear conditioning, as developed by Fanselow (1990). This paradigm has two basic conditions. In one condition, the rats are pre-exposed to the conditioning environment prior to receiving an immediate shock upon being later placed in that environment. In the other condition, rats are only immediately shocked in the environment, with no pre-exposure. If the rat is shocked immediately after being placed in an environment, they fail to show contextual fear. But if the rat had experienced the environment during pre-exposure (without being shocked then), they did show contextual fear after an immediate shock. Presumably, there is minimal exposure period required for the rat to form a conjunctive representation of the context. Pre-exposure allows the rat to form the conjunctive representation even if it is not associated with shock. When the rat is placed in the same environment to be shocked, the features of that environment reactivate the conjunctive representation. Thus, it seems the brief exposure before the shock is sufficient for the reactivation but not the formation of the conjunctive representation.

Rudy and O'Reilly (1999) showed that pre-exposure to the isolated features of a context (e.g., shape of the cage) did not improve contextual fear. Only the presentation of the configuration of features potentiated the fear response in the control rats. Thus, an increase in the salience of the individual features cannot account for the pre-exposure effect. In contrast, a conjunctive representation facilitates the activation of all the features bound by a mere activation of a subset of those features.

Rudy et al. (2002) provided further evidence for the formation of a conjunctive representation during pre-exposure, and its cued recall at the time of immediate shock. This experiment make use of a specific bucket to transport the rat from the housing colony to the pre-exposure environment. Using this same bucket, the rats were then transported to a novel environment and immediately shocked. Fear was then assessed in either the pre-exposure context or the immediate shock one. Control rats showed more fear for the pre-exposure context than the shock one when the old bucket was used. This result is striking because the rats were never shocked in the pre-exposure environment. Thus, the bucket reactivated the pre-exposure context as the rat was taken to the shock environment. When the shock was induced, the rat associated fear with the activated representation. Again, the immediate shock did not allow the formation of a conjunctive representation of the new context. As expected, rats with a hippocampal lesion did not show this effect.

The ability of the bucket to reactivate the memory of the pre-exposure environment provides a direct animal model of cued recall in humans. When a human subject is asked what happened "yesterday," hippocampal pattern completion will activate the representation associated with the word "yesterday." This representation will include the information about what happened and where it happened.

2.2. Comparison with other hippocampal learning theories

The examples above, and a number of others we did not discuss, demonstrate that our computational models can account for a large body of data in both the human and animal literatures. The models are task independent, in the sense that a single common model can be applied to a diverse array of paradigms. Furthermore, the models exhibit differential effects depending on factors such as the similarity of stimuli, type of test effects, etc., which may not be easily summarized with simple dichotomies between memory systems. Nevertheless, existing theoretical dichotomies do capture many of the central tendencies of the model's behavior. In this section, we highlight some of these similarities and differences with existing theories.

2.2.1. Hippocampus: Declarative memory system

Perhaps the most widely accepted theoretical framework is the declarative versus procedural dichotomy (e.g., Squire, 1992). Declarative memory involves remembering facts (e.g., The 21st of December is the shortest day of the year) and events (e.g., I went to the movies yesterday). Procedural memory, on the other hand, involves the acquisition of skills (e.g., learning to play ping-pong) and other forms of non-conscious learning (priming, conditioning, etc.). There is evidence that the media temporal lobe (including the hippocampus and surrounding cortical areas) support declarative memory, while other other cortical and subcortical areas (e.g., basal ganglia and cerebellum) support procedural memory. Overall, we agree that the medial temporal lobe is critical for many aspects of declarative memory. However, we also think that it may play a key role in procedural and other forms of subconscious learning (e.g., Chun & Phelps, 1999), and that other areas play critical roles in declarative memory (e.g., basal ganglia and prefrontal cortex). Thus, this kind of content-based distinction may not map as clearly onto the neural substrates as one based more directly on the neural specializations of the underlying areas, as we have advocated.

2.2.2. Hippocampus: "Representational flexibility"?

Another prominent theory of hippocampal function is centered around the idea that the hippocampus plays a critical role in the acquisition and retention of relational and flexible representations (e.g., Cohen & Eichenbaum, 1993). Specifically, the hippocampus acts as an "associator" of different items and events, and the resulting network of information is not rigidly tied with a specific task but can be accessed and used to support a multitude of goals. This flexible memory system is contrasted with the procedural memory system, which is largely involved in stimulus–response associations. It can also learn stimulus–stimulus associations but this learning cannot be adapted to novel situations. Thus, procedural learning is characterized by the gradual facilitation of a trained association. For example, Eichenbaum, Stewart, and Morris (1990) reported evidence for their theory in a water maze task. They found that both normal and hippocampal rats can learn to locate a hidden platform if they were always released from the same location. When released from a novel location, only the normal rats were successful in finding the platform. They concluded that the hippocampus supports the flexible use of relational information in a novel situation.

Again, many aspects of this theory are consistent with our own. The suggestion that the hippocampus is an associator of individual cues is comparable to our characterization of the hippocampus as the locus of conjunctive representations. In the hidden-platform water maze task, normal rats can locate the platform in a conjunctive representation of all the available distal cues. In contrast, rats with a hippocampal lesion use a response strategy (e.g., turn left) to locate the platform. But when those rats are released from a different location, a left turn may take them away from the platform. However, our model differs from this theory with regard to the role of the hippocampus in behavioral flexibility per se. Indeed, our framework is at odds with this characterization because we maintain that the hippocampus is specialized to learn specific details of events, due to its highly conjunctive, pattern-separated representations. Thus we would not expect the hippocampus to contribute to flexible behavior—by "memorizing"conjunctive features, the hippocampus should treat novel situations as distinct entities, and therefore may actually prevent generalization. Nevertheless, there are specific circumstances in which these hippocampal specializations may indeed be critical for flexibility. To be more clear, we first provide some working criteria for animal behavior to be considered "flexible."

- The ability to flexibly apply or *generalize* acquired knowledge in novel situations.
- The ability to flexibly switch between different behavioral tendencies, depending on the *context* of the environment.
- The ability to flexibly *adapt* to new situations and to change behavior with changing task demands.

Using these criteria, we argue that the term flexibility cannot be assigned a single neural substrate, and further, that specializations of specific brain regions can give rise to flexible behavior in some situations, but they may actually hinder flexibility in others. We discuss this first in terms of hippocampus contributions to flexible behavior, before moving on to the basal ganglia/cortical system which we think is critical for the third kind of flexibility enumerated above.

The ability of the hippocampus to represent spatial context can facilitate flexible behavior in the second sense listed above (context sensitive behavior). The Morris water maze with novel starting location (Eichenbaum et al., 1990) is an example of increased flexibility due to hippocampal representations of the environmental context (i.e., place-field representations). In this case, only the starting point changes, but the context of the environment does not (i.e., it is the same water maze in both cases). Thus, the hippocampus is likely able to pattern complete from the contextual feature stimuli to relevant spatial information, which in turn supports navigation to the hidden platform. Without these hippocampal place-field representations of the overall environment, specific cue–response associations will not generalize to novel starting locations.

However, the context sensitivity imparted by the hippocampus can also lead to less flexible behavior, by preventing the generalization of knowledge (the first sense). For example, there is evidence that the hippocampus can be detrimental to performance in novel situations where spatial information becomes irrelevant. McDonald and White (1994) trained rats to swim towards a visible platform. The rats were later presented with a novel situation: the platform was moved to a new location. Rats with a hippocampal lesion outperformed normal rats by reaching the platform more quickly because they correctly disregarded the old location of the platform. Thus, one might conclude instead that the hippocampus seems to be solely contributing spatial information in the water maze task, instead of some more generalized flexibility capacity.

Similar results were found in a study exploring acquisition of conditioned associations in two different contexts (Honey & Good, 1993). Specifically, rats learned about feature A in one context (A+, context1) and feature B in another (B+, context2). Note that the context is completely irrelevant for acquiring these feature associations. Nevertheless, intact rats exhibited less generalization compared to hippocampally lesioned rats when testing the features in the alternate contexts (A in context2 and B in context1). In this case, it is difficult to say whether the hippocampus is contributing to flexibility (by supporting different behaviors depending on different contexts) or hindering it (by preventing generalization of rewarding behaviors in novel contexts). More generally, we believe that the ability of the hippocampus to rapidly encode novel information is going to be useful in a wide range of different task situations, but that the hippocampus itself is not primarily responsible for the flexible manipulation of the information that it learns. For a detailed example of how hippocampus can play an encoding role, while not supporting flexible retrieval, in the transitive inference task that has been widely cited as supporting the representational flexibility account (e.g., Dusek & Eichenbaum, 1997), see Frank et al., 2003 and associated data VanElzakker, O'Reilly, and Rudy, 2003.

We argue next that the prefrontal cortex/basal ganglia system is specially involved in flexibly selecting adaptive responses in novel situations. Indeed, we argue that the basal ganglia play a dominant role in this function.

3. The basal ganglia: Modulator of cortical representations

As noted earlier, in our model, the basal ganglia (BG) act as a modulatory system that can provide adaptive gating signals to the frontal cortex (e.g., Frank et al., 2001). These gating signals can help to select a particular motor action or larger motor plan from among a number of alternatives currently under consideration. In this way, the basal ganglia contribute to flexible behavior by helping to activate task-appropriate actions, enabling them to overcome prepotent existing associations. This view contrasts in some ways with the prevalent idea that the basal ganglia encode stimulus–response associations as part of the "habit" or procedural learning system. For example, several researchers have found double dissociations in both animals and humans in which the BG are necessary for stimulus–response and procedural learning, whereas the hippocampus is recruited for spatial and episodic memory tasks (e.g., Packard et al., 1989; Packard & Knowlton, 2002; Poldrack et al., 2001; Poldrack, Prabakharan, Seger, & Gabrieli, 1999; Schroeder, Wingard, & Packard, 2002).

In our account of these data, we agree that the hippocampus is important for spatial and episodic memory tasks, whereas the basal ganglia function in tandem with the slower-learning cortical system that cannot rapidly acquire the novel conjunctive representations needed for spatial and episodic tasks. The stimulus–response and procedural learning tasks, on the other hand, require learning to select task-appropriate responses, which is consistent with a modulatory basal ganglia role. We discuss this account in greater detail later, after providing a somewhat more detailed account of our basal ganglia/frontal cortex model.

3.1. The basal ganglia model

If the BG and frontal cortex work together as a system, then what are the respective contributions of the two areas? We argue that different parts of frontal cortex (FC) represent different possible "actions," and that the BG modulate representations in all areas of frontal cortex via distinct anatomical loops (Alexander & Crutcher, 1990; Alexander, Crutcher, & DeLong, 1990). Specifically, we suggest that the role of the BG is to facilitate or suppress actions that are being considered in frontal cortex (Frank, in press; Mink, 1996). In this discussion we will focus on simple motor representations in premotor cortex, but the same arguments can be extended to include cognitive actions, such as the updating of working memory in prefrontal cortex (Frank et al., 2001; O'Reilly & Frank, submitted).

We propose that cortico–cortical connections are involved in selecting multiple possible responses for a given set of incoming sensory stimuli. Without the benefits of a modulator, frontal cortex would try to simultaneously execute all of these responses, leading to high amounts of motor interference. The BG select the most appropriate of these responses by facilitating its execution while suppressing that of competing responses (Mink, 1996). Two main projection pathways from the striatum go through different BG output structures on the way to thalamus and up to cortex, serving the facilitatory and suppressive functions (Figs. 3 and 4). Cells originating in the "direct" pathway inhibit the

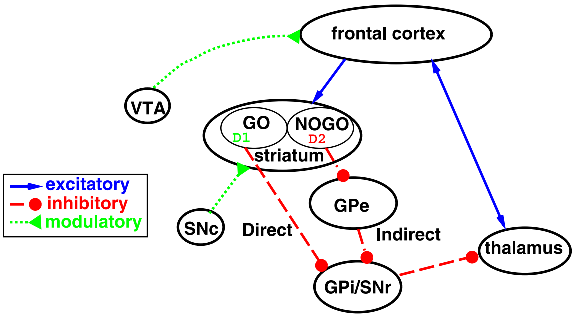


Fig. 3. The cortico–striato–thalamo–cortical loops, including the direct and indirect pathways of the basal ganglia. The cells of the striatum are divided into two sub-classes based on differences in biochemistry and efferent projections. The "Go" cells project directly to the GPi, and have the effect of disinhibiting the thalamus, thereby facilitating the execution of an action represented in cortex. The "NoGo" cells are part of the indirect pathway to the GPi, and have an opposing effect, suppressing actions from getting executed. Dopamine from the SNc projects to the dorsal striatum, differentially modulating activity in the direct and indirect pathways by activating different receptors: The Go cells express the D1 receptor, and the NoGo cells express the D2 receptor. Dopamine from the VTA projects to ventral striatum (not shown) and frontal cortex. GPi, internal segment of globus pallidus; GPe, external segment of globus pallidus; SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; and VTA, ventral tegmental area.

internal segment of the globus pallidus (GPi), whereas the net effect of firing of cells in the "indirect" pathway is to excite the GPi. Because the GPi tonically inhibits the thalamus, direct pathway activity results in thalamic disinhibition, and "gates" the execution of the corresponding command in cortex (Chevalier & Deniau, 1990). Thus, direct pathway activity sends a "Go" signal to cortex, enabling it to execute a given response. Conversely, indirect pathway activity has the opposite effect, sending a "NoGo" signal to suppress competing responses. Note that this disinhibitory interaction with cortex is inherently modulatory in nature, and is very different from a hypothetical alternative where the BG directly excites cortex (Frank et al., 2001).

The above description of BG–FC circuitry is somewhat vague in that it does not specify how the BG "knows" when to signal Go and when to signal NoGo. Our account of how the BG learn this distinction builds on suggestions by Schultz and others that phasic changes in dopamine (DA) firing support learning during reinforcement (e.g, Schultz, 1998, 2002; Schultz, Dayan, & Montague, 1997). Under normal conditions, DA cells fire at intrinsic baseline levels. Unexpected rewards evoke transient bursting of DA cells and increase DA release. By enhancing synaptic plasticity, DA release during unexpected rewards can drive the animal to learn to perform the action that led to the reward Wickens, 1997.

In the BG, the main effect of DA is to enhance Go firing and suppress NoGo firing (Frank, in press). We argue that it is not mere coincidence that Go and NoGo cells primarily express D1 and D2 receptors, respectively

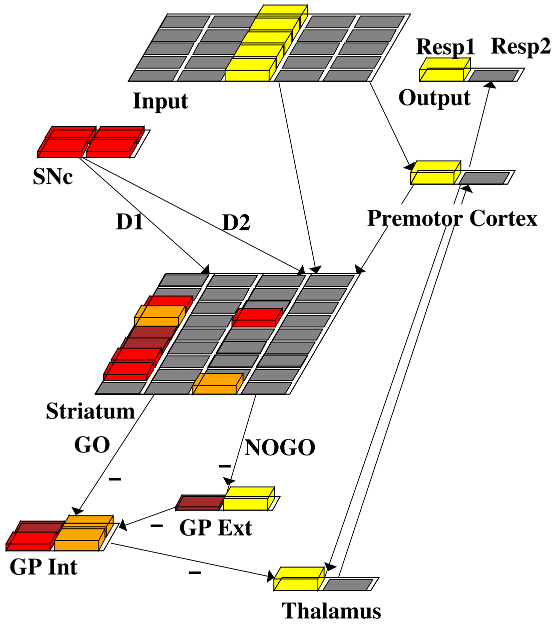


Fig. 4. Neural network model of direct and indirect pathways of the basal ganglia, with differential modulation of these pathways by DA in the SNc. The premotor cortex (PMC) selects a response via direct projections from the Input. BG gating results in bottom-up support from Thalamus, facilitating execution of the response in cortex. In the Striatum, the Go representation for the response (first column) is stronger than its NoGo representation (third column). This results in inhibition of the left column of GPi and disinhibition of the left Thalamus unit, ultimately facilitating the execution of Response1 in PMC. A tonic level of DA is shown here, during the response selection phase. A burst or dip in DA ensues in a second error feedback phase (not shown), depending on whether the response is correct or incorrect for the particular input stimulus. Simulated D1 and D2 receptors are excitatory on the direct/Go pathway and inhibitory on the indirect/NoGo pathway. See Frank (in press) for details.

(Aubert, Ghorayeb, Normand, & Bloch, 2000; Gerfen, 1992). Given that DA is excitatory to synaptic input on D1 receptors (Hernandez-Lopez, Bargas, Surmeier, Reyes, & Galarraga, 1997), its effect is to increase Go activity. And given that it is inhibitory on D2 receptors Hernandez-Lopez et al., 2000, its effect is to suppress NoGo activity. Thus, increases in DA excite Go cells while inhibiting NoGo cells. The resulting increases in Hebbian learning in Go cells may allow the animal to learn to facilitate the action that led to reinforcement.

Note that DA firing can also transiently drop below baseline levels. Indeed, this is consistently observed when animals expect to receive a reward based on previous conditioning, but it is not actually delivered (Hollerman & Schultz, 1998; Satoh, Nakai, Sato, & Kimura, 2003). In this case, NoGo cells may become more excited than their Go counterparts, as they are released from the suppressive influence of DA. Hebbian processes may then enable the animal to learn *not* to subsequently execute the non-reinforcing response, and instead to consider some other response (Frank, in press).

As DA bursts and dips reinforce Go and NoGo representations in the BG, our model showed that the most adaptive (i.e., rewarding) responses represented in premotor areas will tend to get facilitated while less adaptive ones are suppressed. Further, as the BG learn to facilitate adaptive responses, their representations may become enhanced in premotor cortical areas. In this way, DA reward processes within the BG may ingrain prepotent motor "habits" in cortical areas (Frank, in press). Once these habits are ingrained, there is less need for selective facilitation by the BG. This is consistent with observations that dopaminergic integrity within the BG is much more critical for the acquisition than the execution of instrumental responses (Parkinson et al., 2002; Smith-Roe & Kelley, 2000). We have also shown how these same learning principles can shape prefrontal cortex updating signals to solve complex working memory tasks (O'Reilly & Frank, submitted).

In the following sections, we discuss the application of this basal ganglia model within the larger context of our overall tripartite cognitive architecture. First, we discuss the subtle but important distinctions between the functions of action selection versus stimulus response mapping, and how these different views of the BG make different experimental predictions. Next, we argue that the BG may be differentially important for overcoming prepotent response mappings. Then, we return to the issue of behavioral flexibility, discussed above, in the context of more specific BG contributions. Finally, we address the putative direct competition between the HC and BG in place and response learning, which we argue may instead arise from differential modulation of the two structures on motor cortex that follow different time courses.

3.2. Action selection versus stimulus–response mapping

Our model emphasizes the role of the BG in the suppression and selection of cortical representations based on reward history. Those representations include motor responses and plans in the motor and premotor cortices. This characterization is in partial agreement with the stimulus–response (S-R) theory of the dorsal striatal function (for a review, see Packard & Knowlton, 2002). However, our model stresses the involvement of the BG in situations where a number of potent responses are competing to be expressed. In other words, we predict that the BG may not be involved in a situation in which the subject has to learn a single S-R association. In the cued water maze task, for example, rats with a dorsal striatal lesion normally learn to approach a single visible platform (Devan & White, 1999; McDonald & White, 1994). In contrast, the striatal lesion produces a deficit when two visible platforms (one of them does not provide escape) are placed in the maze (Packard & McGaugh, 1992). Thus, the BG seem to be involved when the animal is given a choice between two S-R associations rather than in the association of one stimulus with one response.

Interestingly, a two-choice cue task in a Y-maze does not seem to involve the BG (McDonald & White, 1991; Ragozzino, Ragozzino, Mizumori, & Kesner, 2002b). This is in direct contrast with the results from the two-choice water maze task mentioned above (Packard & McGaugh, 1992). We believe that this discrepancy is also suggestive for a role of the BG in choosing among alternative response options. The Y-maze, unlike the water maze, limits the response options to two: a left or right response. In contrast, the water maze allows an almost unlimited number of possible trajectories. We propose that the BG are needed to enforce the most efficient trajectory towards the stable platform. In fact, rats with dorsal striatal lesion show abnormal swim paths in the watermaze even on the hidden-platform task (Devan, McDonald, & White, 1999; Furtado & Mazurek, 1996; Wishaw, Mittleman, Bunch, & Dunnett, 1987).

A comparison between performance in the Y-maze and the radial maze also led us to the same conclusion. Increasing the number of options in the radial maze (eight options instead of the two in the Y-maze) also reveals a deficit in rats with dorsal striatal lesion (Kantak, Green-Jordan, Valencia, Kremin, & Eichenbaum, 2001; Packard et al., 1989; McDonald & White, 1991; Sakamoto & Okaichi, 2001).

3.2.1. Overcoming prepotent responses

Overcoming a prepotent response is another situation where we predict the involvement of the BG. Such a situation requires the BG to suppress the expression of the prepotent response, and support the selection of a more adaptive one. In fact, the evidence suggests that even when the animal is given two choices (e.g., the Y-maze), a disruption of the BG impairs the performance of reversal and strategy-switching tasks (Ragozzino et al., 2002b; Ragozzino, Jih, & Tzavos, 2002a; Sakamoto & Okaichi, 2001; Wishaw et al., 1987). For example, Ragozzino et al., 2002a reported that the inactivation of the dorsomedial striatum does not affect the acquisition of a response task (i.e., making a left turn in the four-arm plus maze). Yet those rats failed to adjust their behavior in the reversal phase, when they were rewarded for making a right turn. We believe that the BG are involved because reversal creates a difficult choice situation in which the animal has to select a currently rewarded response and suppress a previously rewarded one.

3.2.2. Behavioral flexibility

The basal ganglia/cortical system can contribute to the "adaptive" sense of flexibility as described earlier, by helping to modulate behavior as task demands change. For example, the ability to switch strategies and responses has been widely considered as an example of behavioral flexibility. Despite the term's vagueness, we believe that the use of the term in this context is more fitting than in the description of hippocampal function. Thus, there is evidence that the BG help adaptation to novel situations (such as reversal) irrespective to the content of the modalities involved (Cools, Barker, Sahakian, & Robbins, 2001; Gotham, Brown, & Marsden, 1988; Ragozzino et al., 2002b; Swainson et al., 2000). In one case, Ragozzino et al. (2002b) showed that the BG are essential for switching from a response to a cue strategy and vice versa. In contrast, as we mentioned above, the hippocampus becomes detrimental to performance if a spatial strategy is replaced by a visual cue strategy (McDonald & White, 1994).

In the context of our BG model, we argue that the basal ganglia can learn from negative feedback and help to modulate the execution of motor commands, providing NoGo signals for the no-longer-appropriate actions, and Go signals to the newly appropriate actions. As explained above, this learning depends on the dopaminergic (DA) modulation of Go/NoGo firing in the BG. We have found in our computational models that a sufficient dynamic range of DA signals is required for demanding learning tasks such as in the reversal condition (Frank, in press; O'Reilly, Noelle, Braver, & Cohen, 2002; Rougier, Noelle, Braver, Cohen, & O'Reilly, submitted; Rougier & O'Reilly, 2002). That is, to learn changing reinforcement values of behaviors, the DA signal has to be able to both increase and decrease substantially from baseline levels. Decreases in DA may be necessary not only to suppress initially non-rewarding responses, but may be particularly critical to override responses that were once rewarding but have since changed. This issue was explored in our computational model to explain certain negative effects of dopaminergic medication on cognition in Parkinsons' disease (PD) (Frank, in press).

While medication in PD improves performance in task-switching, it actually tends to impair performance in probabilistic reversal (Cools et al., 2001; Gotham et al., 1988; Swainson et al., 2000). These authors noted that the task-dependent medication effects are likely related to the fact that different tasks recruit different parts of the striatum. Dopaminergic damage in early stage PD is restricted to the dorsal striatum, leaving the ventral striatum with normal levels of DA (Agid et al., 1993; Kish, Shannak, & Hornykiewicz, 1988). This explains why DA medication alleviates deficits in task-switching, which relies on dorsal striatal interactions with dorsolateral prefrontal cortex. However, the amount of medication necessary to replenish the dorsal striatum might "overdose" the ventral striatum with DA, and is therefore detrimental to tasks that recruit it. Thus reversal learning is impaired because it depends on the ventral striatum and ventral prefrontal cortex in monkeys (e.g., Dias, Robbins, & Roberts, 1996; Stern & Passingham, 1995), and recruits these same areas in healthy humans (Cools, Clark, Owen, & Robbins, 2002).

To simulate medication effects, it was hypothesized that medication increases the tonic level of DA, but that this interferes with the natural biological system's ability to dynamically regulate phasic DA changes. Specifically, phasic DA dips during negative feedback may be partially shunted by DA agonists that continue to bind to receptors. When this was simulated in the model, selective deficits were observed during probabilistic reversal, despite equivalent performance in the acquisition phase (Frank, in press), mirroring the results found in medicated patients. Because increased tonic levels of DA suppressed the indirect/NoGo pathway, networks were unable to learn "NoGo" to override the prepotent response learned in the acquisition stage. This account is consistent with similar reversal deficits observed in healthy participants administered an acute dose of bromocriptine, a D2 agonist (Mehta, Swainson, Ogilvie, Sahakian, & Robbins, 2000).

3.3. Place versus response learning

As noted earlier, a number of researchers have hypothesized that the hippocampus and basal ganglia are two competing, parallel learning systems that align essentially with the declarative versus procedural distinction (e.g., Packard et al., 1989; Packard & Knowlton, 2002; Poldrack et al., 1999, 2001; Schroeder et al., 2002). Specifically, the hippocampus is thought to mediate spatial and episodic learning, while the basal ganglia mediates the acquisition of "habits." This view is schematized in Fig. 5A. Our own view of the relationship between BG and hippocampus, based on the computational model described earlier and relevant anatomical projections, is summarized in Fig. 5B. Here, we distinguish between ventral and dorsal striatal areas, which receive preferentially from hippocampus and cortex, respectively (e.g., Groenewegen, Vermeulen-van Der Zee, Te Kortschot, & Witter, 1987). Both BG areas are thought to play a modulatory role on motor responding, but based on different representations encoded by their inputs. Thus, ventral striatum (vBG) modulate responding based on the conjunctive hippocampal representations that provide its input, while dorsal striatum (dBG) provides modulation based on more elemental sensory representations. Indeed, the nucleus accumbens (a principal component of the vBG) is often thought of as the interface between limbic desires and motor output behavior, because it integrates hippocampal and amygdala information to modulate response selection (Mogenson, Jones, & Yim, 1980). The resolution between competing response strategies engendered by these different BG areas and their respective inputs may be resolved within the basal ganglia themselves (e.g., via the

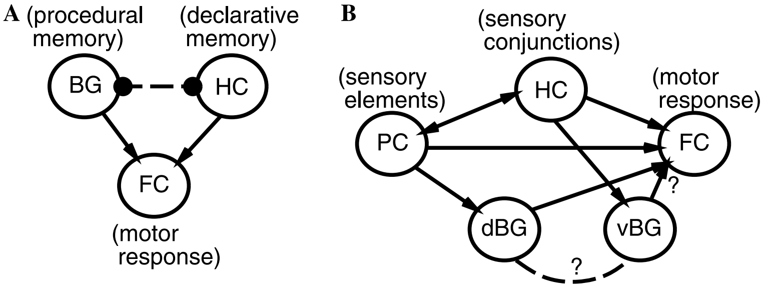


Fig. 5. Two contrasting views of the relationship between basal ganglia (BG) and hippocampus (HC). (A) In one view, the two systems compete directly to drive motor responding, with the HC supporting spatial, place-based behaviors (e.g., "go to this location"), and the BG supporting "habitual" response-based behaviors (e.g., "turn left"). (B) Our view suggests instead that the BG as a whole modulates motor responding based on different kinds of inputs, from HC and posterior cortex (PC). Ventral striatum (vBG) receives from HC, and dorsal striatum (dBG) from PC. Thus, the BG per se does not compete directly with HC, but instead helps to support responding based on its inputs. Similarly, the PC supports responding based on simpler elementary stimuli. Competition between different response strategies may be mediated within the BG itself (e.g., in the globus pallidus) and directly within the frontal cortex response areas.

globus pallidus; Mink, 1996), and also within motor response areas of frontal cortex.

One critical difference between these two views is that we do not view hippocampus as competing with basal ganglia per se. Instead, the hippocampus and the ventral BG work in concert, each contributing specialized functions according to the principles outlined earlier (i.e., hippocampus can rapidly bind information into conjunctive spatial representations, while the vBG can help modulate responding based on these spatial inputs, informed by prior reward-based learning history). These two different views make different predictions and explanations of experimental outcomes.

Some of the relevant data comes from studies where the effects of lesions of the dorsal striatum and hippocampus have been contrasted in the context of a task where either a place-based or response-based strategy can be employed. For example, in a plus maze, rats can be trained to go into a particular arm given the same initial starting arm. This behavior is ambiguous, and could be supported by a response-based strategy (e.g., "go left"), or a place-based strategy (go to this particular location within the overall maze-room environment). Disambiguation comes by placing the rat in the opposite starting arm: a response strategy will cause the rat to go to the opposite location, while a place strategy will result in going to the same location. In such studies, dorsal striatum lesions impair the use of a response-based strategy, while hippocampal lesions impair the use of a place-based strategy (Packard, 1999; Packard & McGaugh, 1996; Poldrack & Packard, 2003). These results have been interpreted as support for a competition between BG and HC. However, it is also consistent with our model, because dBG was lesioned, not vBG. When vBG (specifically nucleus accumbens) is lesioned, however, it reliably produces significant deficits in spatial tasks, similar to those produced by hippocampal damage (e.g., Annett, McGregor, & Robbins, 1989; Roullet, Sargolini, & Mele, 2001; Sargolini, Florian, Oliverio, Mele, & Roullet, 2003; Seamans & Philips, 1994; Setlow & McGaugh, 1998).

Other interesting data from the plus maze and related paradigms show that behavior in intact animals is initially consistent with place responding, but then later transfers to response-based strategies. The fact that hippocampal behavior dominates early in training is consistent with the notion that HC rapidly develops conjunctive representations, while the cortical/BG system slowly ingrains habits.

Other data support the coordinated but still distinct contributions of hippocampus and ventral basal ganglia to spatial processing. For example, Seamans and Philips (1994) deactivated the medial nucleus accumbens with lidocaine injections while rats performed a spatial task in the radial maze. The spatial task involved a sampling phase in which four of the eight arms were baited with food while the other arms were blocked. Thirty minutes later, the rats were required to avoid the four arms visited in the sampling phase because only the remaining arms were baited. Lidocaine injections before the sampling phase had no effect on performance on the test phase. In contrast, injections right before the test phase impaired the rats' ability to avoid the sampled arms.

This study suggests that disabling the nucleus accumbens had no effect on the acquisition of spatial information on the sampling phase. The deficit was found only when the rats were required to respond based on the information acquired on the sampling phase. Thus, it is possible that the medial nucleus accumbens is involved in modulating a response system based on the task demands (win-shift). Sutherland and Rodriguez (1989) reported similar results in the spatial water maze task. They found that rats with a lesion of the whole nucleus accumbens were unable to find a submerged platform using spatial cues. In contrast, a post-acquisition lesion did not affect performance. In other words, the nucleus accumbens seems to be involved in the acquisition but not retention of spatial learning. The results from both experiments suggest that the nucleus accumbens is not involved in either the acquisition of spatial representations nor the navigational performance in a maze. Instead, it may have a time-limited effect in associating spatial representations with a goal-directed strategy (i.e., acquisition of win-shift strategy).

4. Summary and conclusions

To summarize, we have developed a tripartite cognitive architecture based on computational tradeoffs among different types of neural computations that require different parameters and mechanisms. This architecture consists of the posterior cortex, hippocampus, and frontal cortex/basal ganglia system. We have implemented concrete computational models of these different brain areas, and tested their ability to account for a wide range of human and animal behavioral data. In many ways, this computational framework accords well with existing theoretical ideas, but it also makes different predictions in a number of cases. We have argued here that this mechanistic framework may provide a better fit to the data than theories based on verbal dichotomies.

References

Agid, Y., Ruberg, M., Hirsch, E., Raisman-Vozari, R., Vyas, S., Faucheux, B., Michel, P., Kastner, A., Blanchard, V., Damier, P., Villares, J., & Zhang, P. (1993). Are dopaminergic neurons selectively vulnerable to Parkinson's disease? *Advances in Neurology, 60*, 148–164.

Alexander, G., Crutcher, M., & DeLong, M. (1990). Basal ganglia–thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In H. Uylings, C. Van Eden, J. De Bruin, M. Corner, & M. Feenstra (Eds.), *The prefrontal cortex: Its structure, function, and pathology* (pp. 119–146). Amsterdam: Elsevier.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neuroscience, 13*, 266–271.

Anagnostaras, S. G., Maren, S., & Fanselow, M. S. (1999). Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: Within-subjects examination. *Journal of Neuroscience, 19*, 1106.

Annett, L., McGregor, A., & Robbins, T. (1989). The effects of ibotenic acid lesions of the nucleus accumbens on spatial learning and extinction in the rat. *Behavioral Brain Research, 31*, 231–242.

Aubert, I., Ghorayeb, I., Normand, E., & Bloch, B. (2000). Phenotypical characterization of the neurons expressing the D1 and D2 dopamine receptors in the monkey striatum. *Journal of Comparative Neurology, 418*, 22–32.

Burgess, N., & O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus, 6*, 749–762.

Chevalier, G., & Deniau, J. M. (1990). Disinhibition as a basic process in the expression of striatal functions. *Trends in Neurosciences, 13*, 277–280.

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience, 2*(9), 844–847.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Mechanisms of cognitive set flexibility in parkinson's disease. *Brain, 124*, 2503–2512.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22*, 4563–4567.

Devan, B., McDonald, R., & White, N. (1999). Effects of medial and lateral caudate-putamen lesions on place- and cue guided behaviors in the water maze: Relation to thigmotaxis. *Behavioural Brain Research, 100*, 5–14.

Devan, B. D., & White, N. M. (1999). Parallel information processing in the dorsal striatum: Relation to hippocampal function. *Journal of Neuroscience, 19*, 2789.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature, 380*, 69.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America, 94*, 7109–7114.

Eichenbaum, H., Stewart, C., & Morris, R. G. M. (1990). Hippocampal representation in place learning. *Journal of Neuroscience, 10*(11), 3531–3542.

Fanselow, M. S. (1990). Factors governing one-trial contextual conditioning. *Animal Learning and Behavior, 18*, 264–270.

Frank, M.J. (in press). Dynamic dopamine modulation in the basal ganglia: A mjf.da.bib-neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience, 1*, 137–160.

Frank, M. J., Rudy, J. W., & O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus: II. A computational analysis. *Hippocampus, 13*, 341–354.

Furtado, J., & Mazurek, M. (1996). Behavioral characterization of quinolinate-induced lesions of the medial striatum: Relevance for Huntington's disease. *Experimental Neurology, 138*, 158–168.

Gerfen, C. (1992). The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. *Annual Review of Neuroscience, 15*, 285–320.

Gotham, A., Brown, R., & Marsden, C. (1988). 'Frontal' cognitive function in patients with Parkinson's disease 'on' and 'off' levodopa. *Brain, 111*, 299–321.

Groenewegen, H., Vermeulen-van Der Zee, E., Te Kortschot, A., & Witter, M. (1987). Organization of the projections from the subiculum to the ventral striatum in the rat. A study using anterograde transport of *Phaseolus vulgaris* leucoagglutinin. *Neuroscience, 23*, 103–120.

Hasselmo, M. E., & Wyble, B. (1997). Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. *Behavioural Brain Research, 89*, 1–34.

Hernandez-Lopez, S., Bargas, J., Surmeier, D., Reyes, A., & Galarraga, E. (1997). D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca^{2+} conductance. *Journal of Neuroscience, 17*, 3334–3342.

Hernandez-Lopez, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., & Surmeier, D. (2000). D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca^{2+}

currents and excitability via a novel PLCβ1-IP$_3$-calcineurin-signaling cascade. *Journal of Neuroscience, 20*, 8987–8995.

Hikosaka, O. (1998). Neural systems for control of voluntary action—a hypothesis. *Advances in Biophysics, 35*, 81–102.

Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'Reilly, R. C., & Norman, K. A. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus, 12*, 341–351.

Hollerman, J., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1*, 304–309.

Honey, R. C., & Good, M. (1993). Selective hippocampal lesions abolish the contextual specificity of latent inhibition and conditioning. *Behavioral Neuroscience, 107*, 23–33.

Kantak, K. M., Green-Jordan, K., Valencia, E., Kremin, T., & Eichenbaum, H. B. (2001). Cognitive task performance after lidocaine-induced inactivation of different sites within the basolateral amygdala and dorsal striatum. *Behavioral Neuroscience, 115*, 589–601.

Kim, J. J., & Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science, 256*, 675–677.

Kish, S., Shannak, K., & Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. *New England Journal of Medicine, 318*, 876–880.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology, 61*, 260–270.

Mehta, M., Swainson, R., Ogilvie, A., Sahakian, B., & Robbins, T. (2000). Improved short-term spatial memory but impaired reversal learning following the dopamine D2 agonist bromocriptine in human volunteers. *Psychopharmacology, 159*, 10–20.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50*, 381–425.

Mogenson, G., Jones, D., & Yim, C. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology, 14*, 69–87.

Moll, M., & Miikkulainen, R. (1997). Convergence-zone episodic memory: Analysis and simulations. *Neural Networks, 10*, 1017–1036.

Moscovitch, M., & Nadel, L. (1998). Consolidation and the hippocampal complex revisited: In defense of the multiple-trace model. *Current Opinion in Neurobiology, 8*, 297.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychological Review, 110*, 611–646.

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences, 2*(11), 455–462.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). New York: Cambridge University Press.

O'Reilly, R. C., & Frank, M. J. (submitted). Making working memory work: A computational model of learning in the frontal cortex and basal ganglia.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus, 4*(6), 661–682.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

O'Reilly, R. C., Noelle, D., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex, 12*, 246–257.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences, 6*, 505–510.

O'Reilly, R. C., Norman, K. A., & McClelland, J. L. (1998). A hippocampal model of recognition memory. In M. I. Jordan, M. J. Kearns, & S. A. Solla (Eds.), *Advances in neural information processing systems* (Vol. 10, pp. 73–79). Cambridge, MA: MIT Press.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108*, 311–345.

Packard, M. G., & McGaugh, J. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience, 106*, 439–446.

Packard, M. (1999). Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 12881–12886.

Packard, M., & Knowlton, B. (2002). Learning and memory functions of the basal ganglia. *Annual Review in Neuroscience, 25*, 563–593.

Packard, M., & McGaugh, J. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine affects expression of place and response learning. *Neurobiology of Learning and Memory, 65*, 65–72.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience, 9*, 1465–1472.

Parkinson, J., Dalley, J., Cardinal, R., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K., Robbins, T., & Everitt, B. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: Implications for mesoaccumbens dopamine function. *Behavioral Brain Research, 137*, 149–163.

Poldrack, R., & Packard, M. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia, 41*, 245–251.

Poldrack, R., Prabakharan, V., Seger, C., & Gabrieli, J. (1999). Striatal activation during cognitive skill learning. *Neuropsychology, 13*, 564–574.

Poldrack, R. A., Clark, J., PareBlagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 413*, 546–549.

Ragozzino, M., Jih, J., & Tzavos, A. (2002a). Involvement of the dorsomedial striatum in behavioral flexibility: Role of muscarinic cholinergic receptors. *Brain Research, 953*, 205–214.

Ragozzino, M. F., Ragozzino, K. E., Mizumori, S. J. Y., & Kesner, R. P. (2002b). Role of the dorsomedial striatum in behavioral flexibility for response and visual cue discrimination learning. *Behavioral Neuroscience, 116*, 105–115.

R.J., McDonald, & N.M., White (1991). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience, 107*, 3–22.

Rougier, N. P., Noelle, D., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (submitted). Prefrontal cortex and the flexibility of cognitive control: Rules without symbols.

Rougier, N. P., & O'Reilly, R. C. (2002). Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cognitive Science, 26*, 503–520.

Roullet, P., Sargolini, F., & Mele, A. (2001). NMDA and AMPA antagonist infusions into the ventral striatum impair different steps of spatial information processing in a nonassociate task in mice. *Journal of Neuroscience, 21*, 2143–2149.

Rudy, J. W., Barrientos, R. M., & OReilly, R. C. (2002). Hippocampal formation supports conditioning to memory of a context. *Behavioral Neuroscience, 116*, 530–538.

Rudy, J. W., & O'Reilly, R. C. (1999). Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. *Behavioral Neuroscience, 113*, 867–880.

Rudy, J. W., & O'Reilly, R. C. (2001). Conjunctive representations, the hippocampus, and contextual fear conditioning. *Cognitive, Affective, and Behavioral Neuroscience, 1*, 66–82.

Sakamoto, T., & Okaichi, H. (2001). Use of win-stay and win-shift strategies in place and cue tasks by medial caudate putamen (mcpu) lesioned rats. *Neurobiology of Learning and Memory, 76*, 192–208.

Sargolini, F., Florian, C., Oliverio, A., Mele, A., & Roullet, P. (2003). Differential involvement of nmda and ampa receptors within the nucleus accumbens in consolidation of information necessary for place navigation and guidance strategy of mice. *Learning and Memory, 10*, 285–292.

Satoh, T., Nakai, S., Sato, T., & Kimura, M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of Neuroscience, 23*, 9913–9923.

Schroeder, J. P., Wingard, J. C., & Packard, M. G. (2002). Post-training reversible inactivation of hippocampus reveals interference between memory systems. *Hippocampus, 12*, 280–284.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36*, 241–263.

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

Seamans, J., & Philips, A. (1994). Selective memory impairments produced by transient lidocaine-induced lesions of the nucleus accumbens in rats. *Behavioral Neuroscience, 108*, 456–468.

Setlow, B., & McGaugh, J. (1998). Sulpiride infused into the nucleus accumbens impairs memory for spatial water maze training. *Behavioral Neuroscience, 112*, 603–610.

Smith-Roe, S., & Kelley, A. (2000). Coincident activation of NMDA and dopamine D1 receptors within the nucleus accumbens core is required for appetitive instrumental learning. *Journal of Neuroscience, 22*, 7737–7742.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195–231.

Stern, C., & Passingham, R. (1995). The nucleus accumbens in monkeys (*Macaca fascicularis*). *Experimental Brain Research, 106*, 239–247.

Sutherland, R., & Rodriguez, A. (1989). The role of the fornix/fimbria and some related subcortical structures in place learning and memory. *Behavioural Brain research, 32*, 265–277.

Sutherland, R. J., Weisend, M. P., Mumby, D., Astur, R. S., Hanlon, F. M., Koerner, A., Thomas, M. J., Wu, Y., Moses, S. N., Cole, C., Hamilton, D. A., & Hoesing, J. M. (2001). Retrograde amnesia after hippocampal damage: Recent vs. remote memories in two tasks. *Hippocampus, 11*, 27–42.

Swainson, R., Rogers, R. D., Sahakian, B., Summers, B., Polkey, C., & Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia, 38*, 596.

Touretzky, D. S., & Redish, A. D. (1996). A theory of rodent navigation based on interacting representations of space. *Hippocampus, 6*, 247–270.

Treves, A., & Rolls, E. T. (1994). A computational analysis of the role of the hippocampus in memory. *Hippocampus, 4*, 374–392.

VanElzakker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus: I. An empirical analysis. *Hippocampus, 13*, 334–340.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science, 277*, 376–380.

Wickens, J. (1997). Basal ganglia: Structure and computations. *Network: Computation in Neural Systems, 8*, R77–R109.

Wishaw, I., Mittleman, G., Bunch, S., & Dunnett, S. (1987). Impairments in the acquisitation, retention and selection of spatial navigation strategies after medial caudate-putamen lesions in rats. *Behavioural Brain Research, 24*, 125–138.

Wu, X., Baxter, R. A., & Levy, W. B. (1996). Context codes and the effect of noisy learning on a simplified hippocampal CA3 model. *Biological Cybernetics, 74*, 159–165.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.