Multiple memory stores and operant conditioning: A rationale for memory's complexity

Martijn Meeter, Rob Veldkampf, Yaochu Jin

2009

Preprint:

This is an accepted article published in Brain and Cognition. The final authenticated version is available online at: https://doi.org/[DOI not available]

Multiple memory stores and operant conditioning: A rationale for memory's complexity

Martijn Meeter, Rob Veldkamp

Dept. of Cognitive Psychology
VU University Amsterdam, the Netherlands

Yaochu Jin

Honda Research Institute Europe
Offenbach, Germany

Correspondence should be addressed to:
M. Meeter
Dept. of Cognitive Psychology
Vrije Universiteit Amsterdam
Vd Boechorststraat 1
1081 BT Amsterdam
The Netherlands
m@meeter.nl / tel. *31-20-4448993.

Abstract

Why does the brain contain more than one memory system? Genetic algorithms (below referred to with the broader term evolutionary methods) can play a role in elucidating this question. Here, model animals were constructed containing a dorsal striatal layer that controlled actions, and a ventral striatal layer that controlled a dopaminergic learning signal. Both layers could gain access to three modeled memory stores, but such access was penalized as energy expenditure. Model animals were then selected on their fitness in simulated operant conditioning tasks. Results suggest that having access to multiple memory stores and their representations is important in learning to regulate dopamine release, as well as in contextual discrimination. For simple operant conditioning, as well as stimulus discrimination, hippocampal compound representations turned out to suffice, a counterintuitive result given findings that hippocampal lesions tend not to affect performance in such tasks. We argue that there is in fact evidence to support a role for compound representations and the hippocampus in even the simplest conditioning tasks.

Introduction

Being able to learn from experience is crucial for humans and animals. Though it is thus not surprising that sizeable parts of the brain are dedicated to storing information from the past, the sheer complexity of memory in mammals is puzzling. Systems have been identified for storing implicit and procedural memories in the neocortex (Gabrieli, 1998), for habit learning in the striatum (Doya, 2000; Poldrack et al., 2001), for semantic memories in temporal neocortex (Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Murre, Graham, & Hodges, 2001), a retrieval-based episodic memory system has been suggested to be centered around the hippocampus and anterior thalamus (Aggleton & Brown, 1999; Eichenbaum, 1992; Squire, 1992), and a familiarity-based episodic memory system in the parahippocampal gyrus (Aggleton & Brown, 1999). Although there is still debate about individual distinctions, the idea that there are multiple memory systems has become the textbook view.

Why has evolution favored a brain containing more than one memory system? Sherry and Schacter (1987) argued that the concept of *functional incompatibility* might provide an answer. Functional incompatibility arises when the environment poses incompatible demands on a system, and can only be solved by evolving separate subsystems to deal with each challenge. As an example, songbirds tend to store the songs they sing shortly after birth, and retain them throughout the rest of their life. Their memory for songs must thus be very inflexible. Location memory of food storing birds, on the other hand, must be very flexible, so that the bird can update where it stored food and where it took food out. This incompatibility of demands suggests that food-storing songbirds should store songs and food locations in different brain systems, which they indeed do.

Although their line of argumentation is convincing, such reasoning always retains a whiff of a just-so story: If brain architecture had been different, might it have been possible to construct just as plausible evolutionary grounds for that state of affairs? An alternative to such lines of argumentation is to investigate the role of multiple memory systems with in-silico experiments using genetic algorithms (Goldberg, 1989), or more generally, evolutionary algorithms (Fogel, 1995). By simulating variations in architecture and evaluating their survival value, functional grounds can be found for design features such as the multiplicity of memory. There are, however, a number of problems for such an approach. For evolutionary algorithms to work, it must be clear what properties are to be explained through their

contribution to the success or failure of the simulated system. Second, it must be clear how success is measured. What is the task, and in which environment is it carried out? What costs are associated with variations of the system? On all of these accounts, memory is a difficult function to investigate:

- Its exact architecture is imperfectly understood, so it is not clear what properties of memory must be explained.
- It is a central function that plays a role in all domains of behavior. To know the evolutionary benefits of memory, one should thus investigate all possible tasks at once.
- The costs of memory are unclear. How much more expensive, in evolutionary terms, is it to have several memory systems instead of one? Without clear costs evolutionary methods often lead to the rather bland result that high complexity is better than low complexity.

It therefore seems impossible to apply evolutionary methods to analysis of the architecture of memory as a whole. Instead, ways must be found to limit the question to features of memory that are well understood, to a well-known task domain, and to features of brain anatomy of which a cost can be computed. In most episodic memory tasks, performance is not well understood in terms of brain mechanisms. In some simpler task domains, however, brain mechanisms are much clearer (see Klein, Cosmides, Tooby, & Chance, 2002 for a similar argument) This is the case for operant conditioning – learning to initiate some behavior in order to receive a reward. This form of learning is known to rely on the basal ganglia {O'Doherty, 2004 #5393;Yin, 2005 #5425}, and is sufficiently basic to take performance as being of direct survival value.

Moreover, there is a set of properties that can form the basis of selection, and that can be investigated for their survival value. It is known that such simple forms of learning rely themselves on representations of varying complexity (Bouton, Nelson, & Rosas, 1999; Meeter, Myers, & Gluck, 2005; Stanton, 2000). Meeter, Myers and Gluck (2005) have recently proposed that this reliance takes the form of receiving input from different memory stores. Conditioning may occur on the basis of cortical unitary stimulus representations that signals the presence or absence of a grandmother in the input. It may also occur on the basis of ‘episodic’ hippocampal representations, also known as compound or conjunctive representations (Gluck & Bower, 1990; O'Reilly & Rudy, 2000). The

properties that could be investigated for their survival value are the *strengths* or *learning rates* of the connections from different memory stores to the areas that underlie conditioning. Since the metabolic costs of forming and maintaining synapses are relatively high (Attwell & Laughlin, 2001), both strong and rapidly learning synapses carry concrete evolutionary costs (i.e., energy expenditure).

By investigating memory's function as inputs to operant conditioning, we thus have relatively clear survival benefits and costs associated with memory. We can thus use approach the question of why are there different kinds of memory stores in the brain, by translating it into "Are these memory stores all necessary for a naturalistic memory task such as operant conditioning?"

This question will be studied using evolutionary algorithms. Model animals will be given operant conditioning tasks. To perform them, they will have access to different memory stores. Access to any one store implies connections, which is penalized as energy expenditure. The architecture of the animals that will 'win' evolutionary competition will show us which connections, and thus which memory stores, are indispensable for the task, and which can be dispensed with.

Materials and methods

As input to the evolutionary algorithms we used a simple model animal with two systems: a memory system consisting of three stores, and a model basal ganglia that translates inputs from memory into actions. Here we describe the memory system, the basal ganglia model, their integration and the evolutionary simulations. Mathematical details of the memory system model are given in Meeter et al. (2005); details of the basal ganglia model are given in the appendix to this paper.

Model: memory system

We based our model of memory on two earlier models (Meeter et al., 2005; Talamini, Meeter, Murre, Elvevåg, & Goldberg, 2005). In this framework, a variable that distinguishes different memory stores is the kind of representations in those stores. Three stores are modeled.

The Neocortex

The first layer, modeling neocortical processing areas, contains unitary representations of all stimuli presented in the environment. This means that for each stimulus, only a single node will fire

(although in fact cortical neurons are not all that specific in their preferences {e.g., \Gross, 2008 #5424}, it is the case that the stimuli used in conditioning experiments, such as simple sounds and lights, will be processed in different cortical areas). Context information is present in the form of all stimuli that are part of the context, but not as one integrated representation. In this manner, the neocortex provides a distributed representation of the environment.

The Parahippocampus

The second layer, modelling areas in the parahippocampal gyrus, provides a somewhat higher-level representation. As is the case in the brain, model neurons in this area primarily code for identifiable stimuli (Murray, Bussey, Hampton, & Saksida, 2000) but also integrate information from other sources (Witter, Wouterlood, Naber, & Van Haeften, 2000). The parahippocampal layer thus contains *contextually modulated stimulus representations* (Suzuki, Miller, & Desimone, 1997). Another feature of neurons in parahippocampal gyrus (particularly in the perirhinal and entorhinal cortices) is that they respond strongly to novel stimuli, and that this response attenuates with further presentations of the stimulus as if to index stimulus familiarity (Xiang & Brown, 1998). This feature is also included in the model (see Meeter et al., 2005 for details)

The Hippocampus

The third layer models the hippocampus proper. It connects to the parahippocampal layer via broad fanning connections. Diffuse inputs over these connections result in representations not of individual stimuli but of entire situations. Hippocampal representations are often referred to as compound representations; they represent the conjunction of all stimuli present in a situation, but individual stimuli cannot be separated out of the compound. In the model, this means that every new combination of context and a salient stimulus will elicit a new ensemble pattern coding for the configuration, but that no node codes explicitly for stimuli. Indeed, unless stimuli are given behavioral significance, no stimulus representations are found in the hippocampus proper (Young, Otto, Fox, & Eichenbaum, 1997).

Model: basal ganglia

Operant conditioning is a name for learning in paradigms in which environmental feedback shapes operant (i.e., goal-directed) behavior. Since such learning is generally thought to rely on the basal ganglia (e.g., Doya, 2000), a simplified model of the basal ganglia was created based on extant computational work {Brown, 1999 #119; Frank, 2005 #5388}. This model was based on the following, now common, assumptions:

- The dorsal striatum (caudate nucleus and putamen) can release behaviors by lifting inhibition of movement-related centers. It is where responses are learned under influence of rewards {Yin, 2005 #5425}.
- Dopamine signals the presence of rewards, and functions as the learning signal in the dorsal striatum {Schultz, 1986 #4602; Schultz, 2002 #4607}.
- In the regulation of dopamine release the ventral striatum plays an important role {O'Doherty, 2004 #5393;Joel, 2000 #5426}.

In terms of actor critic models (Houk, Adams, & Barto, 1995), the dorsal striatum functions as actor, while the ventral striatum functions as critic (Brown et al., 1999; Houk et al., 1995; O'Doherty et al., 2004). These assumptions are worked out below.

Dorsal striatum

The dorsal striatum was modeled as a layer with one node per possible action. In all simulations reported here, there was just one action to consider – dorsal striatum thus consisted of one node. This node was fully connected with all memory system nodes. If its activation crossed a threshold of 0.5, the action was assumed to be performed.

Since the action had no history of reward at the onset of the simulation, the threshold would not normally be reached in early trials. However, the model can only learn when actions elicit rewards if it performs them. To allow the model to learn when an action was followed by rewards, uniformly distributed noise was added to the activation of the dorsal striatum node in the first half of the trials. If by chance this led to a crossing of threshold in the window in which the action was rewarded, weights from stimulus representations to the node would be strengthened under influence of dopamine. This increased the likelihood of another crossing of threshold in the same situation. At a more abstract

level, the model went through a phase of exploration, in which it learned when to perform the action and when not.

Dopamine

A hypothesis that is gaining general acceptance is that dopamine release plays the role of an *error signal*, signalling an unpredicted reward (Brown et al., 1999; Doya, 2000; Schultz, 2002). Correlations between such an error signal and dopamine release are seen in three dopamine response characteristics. First, when an unexpected reward is delivered to a monkey, dopaminergic neurons in the Substantia Nigra pars compacta (SNpc) and Ventral Tegmental Area (VTA) will engage in burst firing at short latency (Schultz, 1986). Second, when a monkey receives a reward that is reliably predicted by a conditioned stimulus (CS), no change in firing rate of dopaminergic neurons is seen at the moment of the reward, but a small burst is seen at the moment that the CS is presented and the animal knows that a reward is coming (Ljungberg, Apicella, & Schultz, 1991; Schultz, Dayan, & Montague, 1997). Third, when an expected reward is omitted, a *decrease* in firing rate of dopaminergic neurons is seen (Schultz et al., 1997).

Dopamine is known to affect both long-term potentiation (LTP) and long-term depression (LTD) in the dorsal striatum. Dopamine enhances LTP in striatal output neurons via its actions on the D1 receptor (Frey, Matthies, Reymann, & Matthies, 1991; Pawlak & Kerr, 2008). A dip in dopamine levels may enhance LTD in the same neurons (Pawlak & Kerr, 2008). This suggests that dopamine may act as a learning signal. If an action leads to a reward, the associated dopamine release will strengthen the synapses instrumental in producing the action. This may include the synapses from memory systems to the dorsal striatum – it is here that in the model the learning of instrumental responses occurs. If an action ceases to be rewarded, the dip in dopamine release may cause LTD in the synapses that control the action, and may lower the likelihood that the action is performed in the future.

Regulation of dopamine release

There are reasons to believe that the ventral striatum is involved in the regulation of dopamine release from SNpc and VTA. Most prominently, connections are known to exist from both the nucleus accumbens and the ventral pallidum to the VTA and SNpc, either directly or via the pedunculopontine

tegmental nucleus or PPTg (Gerfen, 2004; Yang & Mogenson, 1992). There are several accounts of how dopamine release is controlled; here, we follow the one presented by Brown et al. (1999). A simplified version of their model was implemented.

In their model, dopamine-releasing neurons receive excitatory input via the PPTg from two brain areas: primary rewards are relayed to these neurons via the lateral hypothalamus, while the ventral striatum (mostly nucleus accumbens) relays information about stimuli that predict rewards. Other neuron populations inhibit dopamine release; these neuron populations may learn to predict the timing of an upcoming reward. Timed inhibition from these populations would cancel out excitation from primary reward centers when a reward occurs at a predicted moment. When a reward is predicted but does not occur, the timed inhibition would cause the dip seen in dopamine release. These neural populations are identified by Brown et al. as neurons within the striosomes in the ventral striatum (hereafter referred to as striosomes).

Brown et al. propose the following sequence of events during a standard conditioning experiment in which a CS signals that a reward can be obtained (see Figure 1). Early on in the experiment, the animal will not expect the reward. When a reward occurs, input from the lateral hypothalamus will trigger a burst of firing in dopaminergic neurons. When the animal learns to obtain the reward after each CS (under influence of dopamine; see below), the CS will reliably predict that a reward is coming. Ventral striatal neurons will learn this coincidence, and will excite dopaminergic neurons at the moment that a CS comes in, resulting in dopamine release at that moment. Striosomes will also learn the coincidence. Their activation is timed to the occurrence of the reward, however. At the moment that the reward occurs, striosomes will inhibit dopaminergic neurons, cancelling excitation coming from the lateral hypothalamus. If the reward is then omitted, inhibition from the striosomes is not counteracted by excitation from the lateral hypothalamus, and a dip in dopaminergic neuron firing occurs.

Dopamine bursts will thus occur at the moment of the reward when a reward is not predicted and at the moment of a predicting CS if it is. These dopamine bursts will strengthen LTP in the basal ganglia model. As already stated above, if an animal performs an action that results in a reward, the resultant dopamine burst will strengthen connections from inputs that preceded the action to the model neuron responsible for the action. In other words, dopamine is the teaching signal that tells the model

to what inputs it should emit an action, so that it can earn a reward. Dopamine is also the driver of learning in the ventral striatum and the striosomes.

On a side note, striosomes might also learn to cancel the dopamine burst seen after the predictive CS. This does not occur in vivo. Brown et al. therefore added the assumption that striosomes will only start firing one time step after the CS, leaving time for a short dopamine burst just after the CS. We have taken over this assumption.

Whole model

All three layers of the memory system project to all modeled structures in the basal ganglia. The resultant model is shown in Figure 2. Environmental inputs enter the model via the neocortical layer, while rewards are delivered to the dopamine cell layer in the basal ganglia. The output of the model is an action that is or is not emitted on each time step. For the evolutionary simulations, the model will be understood as a model animal that can perform certain tasks. To allow us to translate the timing of behavioral tasks to model time steps, we tied one time step to a time interval of 150 ms.

The memory system contained 106 nodes and quite a few free parameters. Values for these were taken over from Meeter et al. (2005), however, without alteration. The basal ganglia model contained nine nodes and just two free parameters (the accommodation parameter and the threshold for responding), neither of which influenced behavior of the model very much. What allows the full model to capture different data patterns are the 742 connections between these two parts of the model. With weights on all these connections fixed at 0, the model would not be able to perform. With all connections able to learn, on the other hand, the model can learn essentially arbitrary stimulus-response relationships. What values these connections should have was at the center of the evolutionary simulations.

Evolutionary algorithm

Which memory stores will prove indispensable for operant conditioning tasks? This question was investigated using evolutionary methods. We made use of NSGA-II (Deb, Agrawal, Pratap, & Meyarivan, 2001), which is an efficient multi-objective evolutionary algorithm. Below details are

given of the simulations, including the genotype, fitness values, simulated genetic variations and selection criteria.

Genotype

Since we are interested to see what memory store is important for operant conditioning given a certain task, we need some way of encoding the dependencies on the different memory stores. For this, we chose LTP and LTD learning rates (i.e., the speed with which connections increase and decrease; μ^+ and μ^- in the appendix). These determine the size of connections at the end of the simulation. When evolution develops high LTP learning rates for one of the three memory stores, this will be an indication that operant conditioning is dependent on the representations from that store.

LTD and LTP have opposite effects; a high LTD learning rate would thus indicate a weak dependency to a memory store. This means that many different combinations of LTP and LTD learning rates can have the same results, which makes it hard to analyze the outcome. For that reason we reduced the dimensionality of the problem somewhat by restricting LTD learning rates to only three values. The first value makes the LTD learning rate the same as that of LTP, the second makes it half the LTP learning rate, and the third makes it double the LTP learning rate. LTP rates were given a maximum of 1; not only are very high learning rates physiologically implausible, they also do not contain much information (e.g., in marking the importance of a connection a learning rate of 2 is not much more informative than a learning rate of 1).

Each layer in the basal ganglia model has its own set of three LTP learning rate (one for each of the memory stores). Each memory structure has its own LTD marker. The genotype thus consists of nine LTP learning rates (three for the striosomes, three for the ventral striatum and three for the dorsal striatum) and three LTD markers (one for the neocortex, one for parahippocampus and one for the hippocampus). An example of a genotype is shown in Table 1.

Fitness

There are two objectives for the evolutionary algorithm. One is to minimize the amount of energy used by the model animal in the upkeep of neural connections. The other is to maximize performance, which will be defined shortly, also refer to Equation 1.. Our measure of energy use is simply the sum

of weights between basal ganglia and input structures after learning. This is because weight increases would, in the brain, translate into more and/or stronger synapses, an increase in energy use (Attwell & Laughlin, 2001). The performance measure is more complex. In the paradigms that will be simulated, there are three relevant outcomes: first, the number of rewards obtained (G for *good responses*). This is the number of stimulus presentations followed by the action. Second, the number of wasteful responses; those that were emitted but not rewarded (B for *bad responses*). The third are the number of potential rewards that were missed, equal to the number of stimulus presentations that were not followed by the action. These were taken together in one formula:

Equation 1 $$P(G, B, R) = \frac{G - \alpha B}{R}$$

Here, P stands for performance, and *R* for the number of potential rewards. The number of missed rewards is implicit in this formula as the gap between *G* and *R*. The penalty given for wasteful responses can be set by changing α (here set to 0.01). It is easy to see that the performance value will be 1 for creatures responding to every opportunity without ever making an unnecessary or incorrect response. The responses emitted during exploration will generally lower performance, as actions will during that phase also be emitted when they are not rewarded. Therefore, performance will never be equal to 1.0.

Genetic Variations and Selection

The evolution starts with an initial parent population consisting of 50 model animals whose genotype is randomly generated. In each generation of the evolution, parent individuals are varied to create offspring individuals. In this work, simulated binary crossover (SBX) developed by Deb and Agawal (1995) and polynomial mutation (Deb & Goyal, 1996) are adopted. In SBX, two offspring are created by sampling from a probability distribution based on the locations of the two parents and a distribution index (η_c). The smaller η_c , the wider the distribution will be and the more explorative the evolutionary search will be. Polynomial mutation varies the offspring based on a polynomial probability distribution defined again by a distribution index (η_m). In our experiments, both distribution indexes are set to 20. During reproduction, two parents are chosen from the parent population using binary

tournament selection. The two parents are then crossed over using the SBX with a probability of 0.9 to generate two new solutions, which are further mutated using polynomial mutation with a probability of 0.16. This procedure is repeated 25 times to generate 50 offspring. All 50 individuals go through the simulated operant conditioning task (reported below), so that the energy efficiency and performance of each model animal can be evaluated. Afterwards, the crowded non-dominated sorting algorithm (Deb et al., 2001) is applied to select 50 individuals from a combination of parent and offspring populations according to individuals' fitness values (energy consumption and performance) as well as the crowdedness in the individuals' neighborhood. The crowdedness is measured by the sum of the Euclidean distances between the individual and its two neighbors, which is used as a criterion in selection to ensure that the population is able to find diverse Pareto-efficient solutions. Each simulation lasted for 99 generations. This number was chosen as by 99 generations the population was stable (this was checked in the most complex, third simulation, where 199 generations did not lead to different results). At the end of evolution (i.e., after 99 generations), all Pareto-efficient model animals (for which we will use the term *solutions*) were analyzed.

Simulated tasks

Simple operant conditioning

The first experiment analyzed a simple operant conditioning paradigm. In a simulated environment, a stimulus was presented every 7.5 seconds (50 time steps) for 150 ms. A response after 300 but within 750 ms of the CS (3 to 5 time steps) was rewarded. Outside this window, rewards were not assigned. The stimulus was presented for 40 trials.

Stimulus discrimination

In a second experiment, two stimuli were presented in alternating blocks of 40 trials. A response following the first stimulus was rewarded in the same way as in the simple operant conditioning simulation. After the second it was not rewarded. This second stimulus was presented for 40 trials> Lastly, the first stimulus was again presented for 40 trials.

Context discrimination

In a third experiment, one stimulus was presented in two contexts. A response after the CS was rewarded in the first context, but not in the second. The CS was first presented for 40 trials in the first context, then for 40 trials in the second, then for 40 trials in the first.

Results

All three experiments yielded multiple solutions that were Pareto efficient. There were few solutions in the context discrimination experiment; for that reason we ran it twice. Performance and energy of the solutions found in all thee experiments are given in Figure 3. Performance was generally highest in standard operant conditioning. Solutions in the context discrimination experiment varied more than in the other experiments, where they tightly clustered together.

Turning to the underlying genotypes, Table 2 gives average LTP rates for solutions in the three experiments. LTD rates were not given; they tended to be high for high LTP rates and vary more strongly for low LTP rates (i.e., if the LTP rate for a connection was high, the LTD rate was almost always double the LTP rate, while all three LTD multiples could appear for low LTP rates).

In the standard operant conditioning experiment, all connections to the ventral striatum and the striosomes had high LTP rates. This means that the three kinds of representations were all needed for adequate control of the dopamine signal. The same was true in the CS discrimination experiment and, to a lesser extend, in the context discrimination experiment.

LTP rates on the connections to the dorsal striatum varied much more strongly. Since the connections to the dorsal striatum ultimately determine responses, these results are presented in more detail in Figure 4, in which both the average and the range are given of LTP rates to the dorsal striatum in the solutions. In the standard operant conditioning experiment, LTP rates were high on the connections from the parahippocampal and hippocampal layers to the dorsal striatum. From the cortical layer, they were negligible, suggesting that this input was not important. The same was true in the CS discrimination experiment, in which only LTP rates on the connection from the hippocampal layer diverged from zero. In the context discrimination experiment, on the other hand, LTP rates on

connections from all three layers were above zero. This suggests that context discrimination, but not CS discrimination or learning about a single CS, necessitates input from all three layers.

Discussion

Here, we investigated the role of multiple memory stores in operant conditioning as a proxy for evolutionary survival value. To do this, we implemented a simple model of the basal ganglia, with a ventral striatal and striosomes layer that together control dopamine release (functioning as critic in terms of actor-critic models), and a dorsal striatum layer that controls the release of actions (functioning as actor). With this model, we simulated paradigms of operant conditioning. Though our model reflects current thinking about the roles of the basal ganglia and dopamine in learning {e.g., \O'Doherty, 2004 #5393;Yin, 2005 #5425;Brown, 1999 #119; Frank, 2005 #5388;Schultz, 1986 #4602; Schultz, 2002 #4607}, it could still be an incorrect account of the brain substrate of conditioning. This would invalidate our results. All conclusions drawn below are thus contingent on that our account of operant conditioning is correct.

A key distinction between memory stores, and one we chose to focus on, is between cortical unitary stimulus representations and hippocampal compound representations in which the stimulus and its context are packed together. Other dichotomies have also been attached to cortical and hippocampal stores, such as that between gradual learning of habits, knowledge and skills, and one-shot learning of episodes (Kinsbourne & Wood, 1975; McClelland, McNaughton, & O'Reilly, 1995; Meeter & Murre, 2005; Sherry & Schacter, 1987). We have not investigated these, but they were implicitly present in our use of fixed, localized representations in the cortical layer and self-organizing distributed representations in the hippocampal layer.

In simple conditioning to a stimulus, conditioning can occur either on the basis of cortical unitary stimulus or of hippocampal compound representations. Results suggest that both unitary and compound representations were necessary inputs to the ventral striatal and striosomes layers that control dopamine release. That was not true for inputs to the lateral striatal layer, the layer that controls actions. Here, only context discrimination required multiple inputs. For simple operant conditioning and CS discrimination, compound representations as found in the hippocampal layer sufficed. This suggests that hippocampal compound representations may be used by default in

conditioning tasks such as studied here, while neocortical unitary representations may play a role in more complex forms of learning.

At first blush the results seem both consistent and inconsistent with established findings from neuroanatomy and from lesion studies. Focusing first on neuroanatomy, the ventral striatum in mammals receives dense projections from both hippocampal region structures and from neocortical structures (Gerfen, 2004), as predicted by our results. The dorsal striatum, on the other hand, receives most inputs from cortical structures (Gerfen, 2004), whereas our simulations suggested that hippocampal representations are also a basis of dorsal striatal functioning. However, the most important cortical input to the dorsal striatum comes from prefrontal cortical areas, which themselves receive a dense hippocampal innervation (Witter & Amaral, 2004).

With regard to lesion studies, such studies support our result that both hippocampal region and cortical areas are important in contextual discrimination paradigms (e.g., Fanselow, 2000; Kim & Baxter, 2001). However, some of the same studies also suggest that hippocampal region areas are not involved in paradigms in which stimuli must be discriminated (e.g., Fanselow, 2000; Kim & Baxter, 2001). Moreover, in classical conditioning hippocampal lesions leave the ability to learn to respond to individual stimuli largely unperturbed (Schmaltz & Theios, 1972). This is counter to our results, which suggested that the hippocampus is a main input when stimuli need to be discriminated.

A closer look at the literature, however, suggests that there is no real contradiction between our results and previous ones. Lesion studies show what brain regions are necessary for a certain behavior. Evolutionary simulations, on the other hand, can show what regions would be sufficient for that behavior. Lesion studies show that the hippocampus is not necessary for stimulus discrimination, whereas our results suggest that it may be sufficient for such tasks. What gives these findings an air of contradiction is the intuition that if a brain region is not necessary for a certain behavior, it does not underlie that behavior in intact animals. In the case of simple conditioning to stimuli and the hippocampus, several findings suggest that this intuition is wrong: the hippocampus does seem to have a strong role in learning about stimuli. For example, intact rats that have been conditioned to stimulus A will stop responding when this stimulus is suddenly accompanied by a stimulus B. This is not seen in hippocampectomized rats (Allen, Padilla, Myers, & Gluck, 2002), suggesting that in the intact rats the hippocampus is involved in the response to stimulus A. Moreover, responses to simple stimuli in

classical conditioning are disrupted by a contextual change early in learning but not late in learning. This is not seen in hippocampectomized animals (Myers & Gluck, 1994). Again, this suggests that responses, at least early in learning, are based on hippocampal representations in intact animals. At closer inspection, findings from the hippocampal lesion literature thus corroborate, rather than refute, our results.

Why then, would the hippocampus play a role in simple conditioning, even though lesion data suggest it is not necessary in such tasks? There may be two reasons for this. The first is that compound representations deliver a flexibility that is lacking in unitary representations. When an animal must learn to predict a reward, it cannot in advance know whether the reward is tied to a stimulus or to a context, or to a combination of the two. As a consequence, it may be advantageous to use flexible memory representations. To give an ecologically inspired example, if a scavenger finds a carcass, it cannot know in advance whether future finds of the same kind are predicted by the location (i.e., it is a place in which diseased animals retreat), by a sound (e.g., a lion's roar), or whether only a combination of a sound and a place will predict a similar find. Such flexibility may play an important role in explaining experimental findings, but not in our results as our model animals received only one task and so had no need for flexibility.

The second reason, and the one that explains our results, is one of representational economy. The task in the operant conditioning paradigms simulated here is to differentiate situations in which an action will be rewarding (i.e., when a stimulus is being presented or just has been presented) from situations in which the action is not rewarded (i.e., when only the context is present). Within cortical areas all inputs that impinge on the senses of the animal elicit some firing. Stimuli used in conditioning tend to have discreet onsets and offsets, and are therefore likely to capture attention and receive preferential processing (Bundesen, Habekost, & Kyllingsbaek, 2005; Theeuwes, 1994). Nevertheless, it is likely that of all spikes within neocortical areas, only a minority is related to a stimulus. This is certainly true within the model. Only a minority of the firing can thus be used to discriminate situations in which the action is rewarding from those in which it is non-rewarding. Within the hippocampus, on the other hand, all spikes are related to the situation at hand, which in the model is the ensemble of stimulus and context. This means that differentiating situations that predict

reward (i.e. stimulus + context) from those that do not predict reward (context alone) can occur economically.

Findings from comparative neurology support the view of hippocampal compound representations as default. First, a hippocampal region is present in all mammals, and seems to have roughly the same role in all. Moreover, as a percentage of brain volume, the hippocampal region is bigger in smaller mammals than in bigger ones (Stephan, 1983). Moreover, even in insects compound representations can be found, and can form the basis of learning in conditioning paradigms. An example is the honeybee, whose mushroom bodies contain compound representations (Menzel, 2001). These are used in simple conditioning to odors, as is evident from the context sensitivity of such learning in its early phase (Menzel, 2001).

In summary, our results suggest that having multiple representations is important in learning to regulate dopamine release, as well as in learning to discriminate in which contexts a stimulus signals the opportunity for reward. For simple operant conditioning, as well as learning to discriminate two stimuli, one kind of representations sufficed. Hippocampal compound representations turned out to be an ideal basis for such learning, a counterintuitive result given anatomical considerations and standard interpretations of lesion studies. We argue, however, that at closer reading the literature actually supports a role for compound representations and the hippocampus in simple conditioning.

Our work also leads to testable predictions. In intact animals, instrumental conditioning should be sensitive to context change in early phases of learning, just as in classical conditioning (Myers & Gluck, 1994). Moreover, contrary to general expectations, eliminating neocortical inputs from the basal ganglia should affect contextual discriminations more than simple discriminations. This prediction cannot be tested with neocortical lesions as these would also deafferentiate the hippocampal lobe. A selective lesion of connections from neocortical regions coding for one modality to the dorsal striatum might be possible, however, though no such lesion technique has been reported as yet.

Literature

Aggleton, J.P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences, 22*, 425-489.

Allen, M. T., Padilla, Y., Myers, C. E., & Gluck, M. A. (2002). Selective hippocampal lesions disrupt a novel cue effect but fail to eliminate blocking in rabbit eyeblink conditioning. *Cognitive, Affective, and Behavioral Neuroscience, 2*, 318-328.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism, 21*, 1133-1145.

Bouton, M. E., Nelson, J. B., & Rosas, J. M. (1999). Stimulus generalization, context change, and forgetting. *Psychological Bulletin, 125*, 171-186.

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience, 19*, 10502-10511.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112*, 291-328.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A semantic subsystem of visual attributes. *Neurocase, 4*, 499-505.

Deb, K., & Agrawal, R. B. (1995). Simulated binary crossover for continuous search space. *Complex Systems, 9*, 115-148.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2001). A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*, 182-197.

Deb, K., & Goyal, M. (1996). A combined genetic adaptive search (GeneAS) for engineering design. *Computer Science and Informatics, 26*, 30-45.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology, 10*, 732-739.

Eichenbaum, H. (1992). The hippocampal system and declarative memory in animals. *Journal of Cognitive Neuroscience, 4*, 217-231.

Fanselow, M. S. (2000). Contextual fear, gestalt memories, and the hippocampus. *Behavioural Brain Research, 110*, 73-81.

Fogel, D. B. (1995). *Evolutionary Computation - Toward a New Philosophy of Machine Intelligence*. New York: IEEE Press.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience, 17*, 51-72.

Frey, U., Matthies, H., Reymann, K. G., & Matthies, H. (1991). The effect of dopaminergic D1 receptor blockade during tetanization on the expression of long-term potentiation in the rat CA1 region in vitro. *Neuroscience Letters, 129*, 111-114.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*, 87-115.

Gerfen, C. R. (2004). Basal ganglia. In G. Paxinos (Ed.), *The rat nervous system, 3rd ed.* (pp. 445-508). Amsterdam: Elsevier Academic Publishers.

Gluck, M. A., & Bower, G. (1990). Component and pattern information in adaptive networks. *Journal of Experimental Psychology: General, 119*(1), 105-109.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison-Wesley.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia*. Cambridge, MA: MIT Press.

Kim, J. J., & Baxter, M. (2001). Multiple brain memory systems: The whole does not equal the sum of its parts. *Trends in Neuroscience, 24*, 324-330.

Kinsbourne, F. A., & Wood, D. (1975). Short-term memory processes and the amnesic syndrome. In D. Deutsch & J. A. Deutsch (Eds.), *Short-term memory*. New York: Academic press.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review, 109*, 306-329.

Ljungberg, T., Apicella, P., & Schultz, W. (1991). Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Research, 567*, 337-341.

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.

Meeter, M., & Murre, J. M. J. (2005). TraceLink: A connectionist model of consolidation and amnesia. *Cognitive Neuropsychology, 22*(559-587).

Meeter, M., Myers, C. E., & Gluck, M. A. (2005). Integrating incremental learning and episodic memory models of the hippocampal region. *Psychological Review, 112*, 560-585.

Menzel, R. (2001). Searching for the memory trace in a mini-brain, the honeybee. *Learning & Memory, 8*, 53-62.

Murray, E. A., Bussey, T. J., Hampton, R. R., & Saksida, L. M. (2000). The parahippocampal region and object identification. *Annals of the New York Academy of Sciences, 911*, 166-174.

Murre, J. M. J., Graham, K. S., & Hodges, J. R. (2001). Semantic dementia: Relevance to connectionist models of long-term memory. *Brain, 124*, 647-675.

Myers, C. E., & Gluck, M. A. (1994). Context, conditioning and hippocampal re-representation. *Behavioral Neuroscience, 108*, 835-847.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. J., & Dolan, R. (2004). Dissociable roles of dorsal and ventral striatum in instrumental conditioning. *Science, 304*, 452–454.

O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus, 10*, 389-397.

Pawlak, V., & Kerr, J. N. (2008). Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *Journal of Neuroscience, 28*, 2435-2446.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C. E., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546-550.

Schmaltz, L., & Theios, J. (1972). Acquisition and extinction of a classically conditioned response in hippocampectomized rabbits (Oryctolagus cuniculus). *Journal of Comparative and Physiological Psychology, 79*, 328-333.

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of Neurophysiology, 56*(5), 1439-1461.

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*, 1593-1599.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review, 94*, 439-454.

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

Stanton, M. E. (2000). Multiple memory systems, development and conditioning. *Behavioural Brain Research, 110*, 25-37.

Stephan, H. (1983). Evolutionary trends in limbic structures. *Neuroscience and Biobehavioral Reviews, 7*, 367-374.

Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *Journal of Neurophysiology, 78*(2), 1062-1081.

Talamini, L. M., Meeter, M., Murre, J. M. J., Elvevåg, B., & Goldberg, T. E. (2005). Reduced parahippocampal connectivity produces schizophrenia-like deficits in simulated neural circuits with reduced parahippocampal connectivity. *Archives of General Psychiatry, 62*, 485-493.

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception, 23*(4), 429-440.

Witter, M. P., & Amaral, D. G. (2004). Hippocampal formation. In G. Paxinos (Ed.), *The rat nervous system* (3rd ed., pp. 637-703). Amsterdam: Elsevier.

Witter, M. P., Wouterlood, F. G., Naber, P. A., & Van Haeften, T. (2000). Anatomical organization of the parahippocampal-hippocampal network. *Annals of the New York Academy of Sciences, 911*, 1-24.

Xiang, J. Z., & Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity, and recency in regions of the anterior temporal lobe. *Neuropharmacology, 37*, 657-676.

Yang, C. R., & Mogenson, G. J. (1992). Hippocampal signal transmission to the pedunculopontine nucleus and its regulation by dopamine D2 receptors in the nucleus accumbens: an electrophysiological and behavioural study. *Journal of Comparative Neurology, 323*, 387-410.

Young, B. J., Otto, T., Fox, G. D., & Eichenbaum, H. (1997). Memory representation within the parahippocampal region. *Journal of Neuroscience, 17*(13), 5183-5195.

Appendix: details of the basal ganglia model

Here, a formal description is given of the basal ganglia model described verbally in the main text. We use the letters M, D, V, S and P to denote the output of the memory, dorsal striatum, ventral striatum, striosomes and PPTg layers respectively. All of these layers contain only a single node, with the exception of the striosomes layer that contains five nodes. The letter X will be used whenever multiple layers are governed by the same formula, and the letter W^X for weights on the connections from some layer *X*. Indexes for time steps will only be given if outputs from different time steps are used in the calculations. On every time step all activations are calculated first; weights are updated thereafter. We will explain the computations in the order of calculation.

As a first step, activations in the memory system are computed (see Meeter et al., 2005 for details). Then, the input vector *M* to the ventral and dorsal striatal nodes is obtained by taking, for each node *m* in the three memory system layers, the average output over the last 5 time steps. The activation of striatal node *i* in layer *X* (dorsal or ventral striatum) is then computed as:

$$A_i^X = \left(1 - G_i^X\right) \sum_m W_{i,m}^X M_m^5 + \delta \qquad (1)$$

Here, M^5_m stands for the input from memory node *m* averaged over the current and 4 previous time steps. G^X_i gives the accommodation of node *i* (see equation 5), and $W^X_{i,m}$ gives the weight on the connection between node *m* and node *i*. δ is added noise. No noise is added to the ventral striatal node. Dorsal striatal neurons receive uniformly distributed noise with range [0, 1] during the first half of the experiment to implement exploration.

In the striosomes layer, each of the five nodes integrates input over a different number of time steps: the first only takes input from the current time step, the second from the current and last time step, etc.

$$A_i^S = \left(1 - G_i^S\right) \sum_m W_{i,m}^S M_m^i \qquad (2)$$

Here, M'_m stands for the input from memory node *m* averaged over the current and previous *i-1* time steps.

Activation of the PPTg node is a function of its accommodation, of primary reward and of the ventral striatal input to it:

$$A^P = (1 - G^P)(R + \sum_i V_i) \quad (3)$$

Here, *R* is primary reward, delivered to the model on the time step after it performed a rewarded action.

Output X_i of node *i* in all layers *X* except the dorsal striatum is equal to activation A^X_i if that activation is between 0 and 1, else it is equal to either 0 or 1:

$$X_i = A_i^X \text{ if } A_i^X \text{ between 0 and 1} \quad (4a)$$

$$X_i = 0 \text{ if } A_i^X < 0 \quad (4b)$$

$$X_i = 1 \text{ if } A_i^X > 1 \quad (4c)$$

Output D_i of the dorsal striatum is 1 if A_i^D is larger than 0.22 and 0 otherwise. On all time steps that D_i is 1, action *i* is performed by the model animal. If the action falls within a reward period, *R* is set to 1 on the next time step.

Accommodation is computed in the same way for all nodes *i* in the ventral and dorsal striatal, PPTg and striosomes layers. Using index *t* for the current time step, accommodation of the next time step *t+1* is based on accommodation and output in the current time step *t*:

$$G_i^X(t+1) = gG_i^X(t) + X_i \quad (5)$$

The accommodation constant *g* is set to 0.15, which implies a quick rebound out of accommodation.

Dopamine output *dop* is calculated as the difference between excitatory PPTg input and inhibitory input from the striosomes:

$$dop = P - \sum_i S_i \qquad (6)$$

dop represents the level of dopamine outflow from nigral and VTA dopaminergic cells relative to baseline. It can thus also become negative, which would model a dip in dopamine outflow. It influences learning in the model, but does so differently when positive and when negative. Therefore in the learning rule two variables are used: dop^+, which is equal to *dop* if *dop* > 0 (and 0 otherwise), and dop^-, which is equal to *dop* if *dop* < 0 (and 0 otherwise). The learning rule is, for dorsal and ventral striatal nodes and nodes in the striosomes layer:

$$\Delta W^X_{i,m} = X_i \left[\mu^+ \overline{M}^n_m \left(1 - W^X_{i,m}\right) dop^+ - \mu^- \left(1 - \overline{M}^n_m\right) W^X_{i,m} dop^- \right] \qquad (7)$$

This variant of Hebbian learning makes the change in the weight from node *m* to node *i*, both positive and negative, a function of the strength of the input from node *m* and of the output of node *i*. Again, input is averaged over a number of time steps – that number of time steps, here denoted by *n*, is the same as in equations 1 and 2. Weights are increased when dop^+ is larger than 0, and decreased when dop^- is larger than 0 (i.e., when *dop* is positive or negative respectively). Weight increase is large when inputs are strong, and when a weight is far from its maximum value of 1. Weight decrease is large when inputs are weak, and when the weight is far from its minimum value of 0.

Tables

Table 1. Example of a genotype. Each value in the first three rows shows the LTP learning rate between a memory store and a basal ganglia structure. The fourth row shows the marker for LTD. A value of 0.5 means that LTD learning rate is half the LTP learning rate. For example, the LTD learning rate from the parahippocampal layer to the ventral striatum would in the example be 0.32 * 0.5 = 0.16. The marker can also be 1 (equal learning rates) or 2 (LTD twice as strong as LTP).

	Neocortex	Parahippocampus	Hippocampus
Dorsal striatum	0.23	0.02	0.01
Ventral Striatum	0.04	0.32	0.03
Striosomes	0.24	0.34	0.01
LTD marker	1	0.5	2

Table 2: For each experiment and each connection in the model, the average LTP rate in the solutions (e.g., the upper left number represents the average LTP rate on the connection from the cortical layer to the ventral striatum layer, in the model animals that were solutions in the standard operant conditioning experiment).

	to ventral striatum			to striosomes			to dorsal striatum		
from	cortex	parahip	hippoc	cortex	parahip	hippoc	cortex	parahip	hippoc
standard	0.43	0.55	0.76	0.88	0.99	0.98	0.00	0.92	0.92
CS discrim.	0.06	0.33	0.59	0.86	0.75	0.99	0.00	0.01	0.64
context discrim.	0.47	0.56	0.58	0.30	0.72	0.71	0.46	0.47	0.53

Figure 1: Characteristics of dopaminergic neuron firing in three situations, and schematic of the account of the Brown et al. (1999) model. Each primary reward elicits a burst of dopaminergic neuron firing via inputs from the lateral hypothalamus (via the PPTg). Conditioned stimuli (CS) activate representations in cortical memory systems. These in turn project to the ventral striatum and to striosomes. If a CS predicts a reward, this relation is stored in both areas. The ventral striatum will fire at the time of the CS, eliciting dopaminergic firing at the time of the CS. Striosomes will fire at the time of the predicted reward, and inhibit dopaminergic firing. Their inhibitory input will cancel out excitation from the lateral hypothalamus if a well-predicted reward occurs, and will cause a dip in dopaminergic firing if a well-predicted reward is omitted. Recordings of dopamine cells in monkeys taken from Schultz et al. (1997).

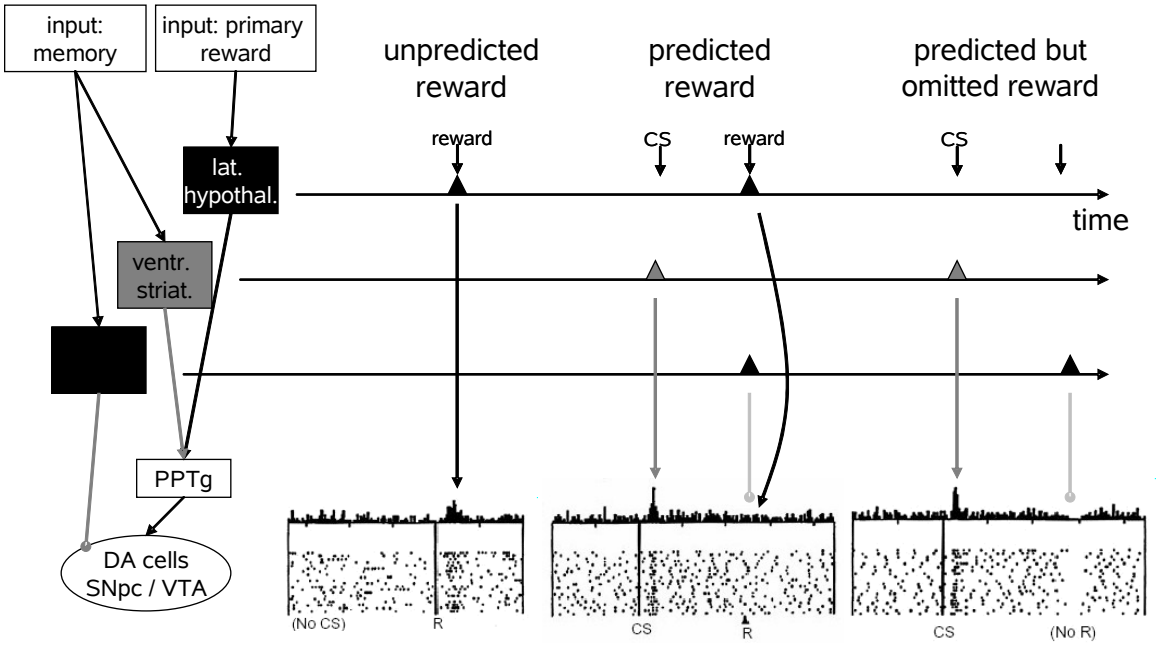

Figure 2: Whole model used in the simulations. The memory system consists of three layers that form different representations of the input, with the input consisting of context (all stimuli that are continuously present) and CS'ses, stimuli with discrete presentation times. The basal ganglia model consists of a dorsal striatum layer that codes for actions, a ventral striatum layer that learns with stimuli predict rewards, a striosomes layer that learns to inhibit dopamine output, a PPTg node that relays signals to dopminergic cells, and a dopamine cell layer that gives off dopamine to the other basal ganglia structures. Actions can produce rewards, which via the lateral hypothalamus (not modeled) elicit dopamine firing.

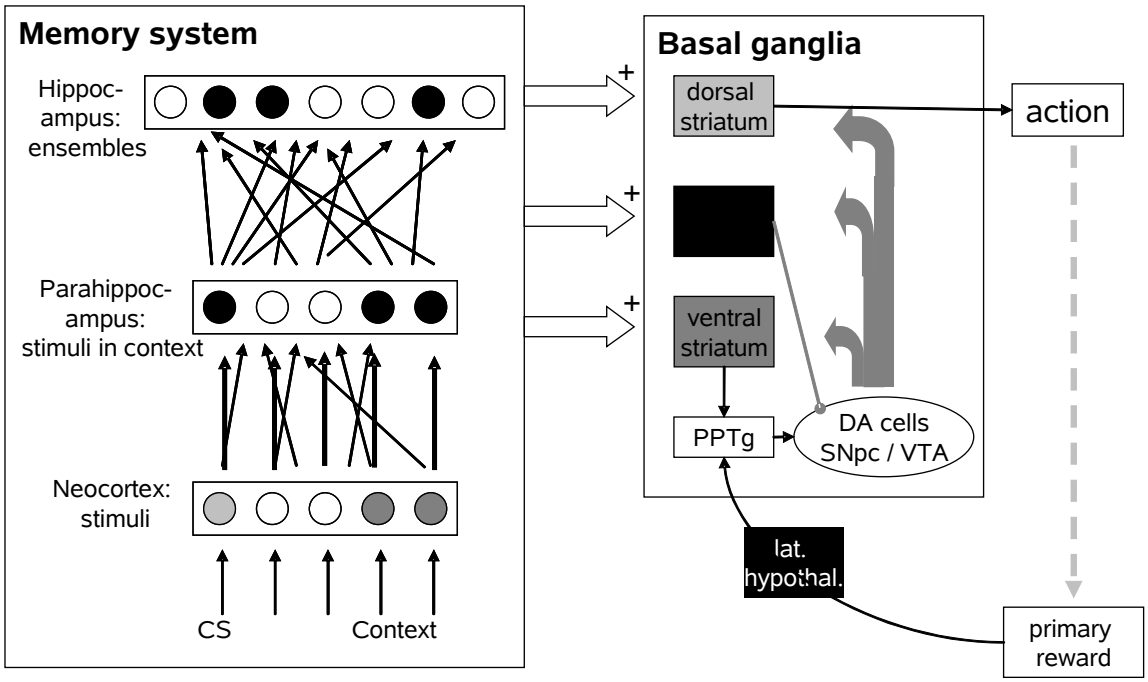

Figure 3: Energy (measured as summed weights at the end of the simulation) and performance (as a function of the number of received rewards and the number of unnecessary responses) for all solutions in the three experiments: standard operant conditioning, CS discrimination, and context discrimination. The scale of the energy measure is inverted as lower energy translates into higher fitness for model animals.

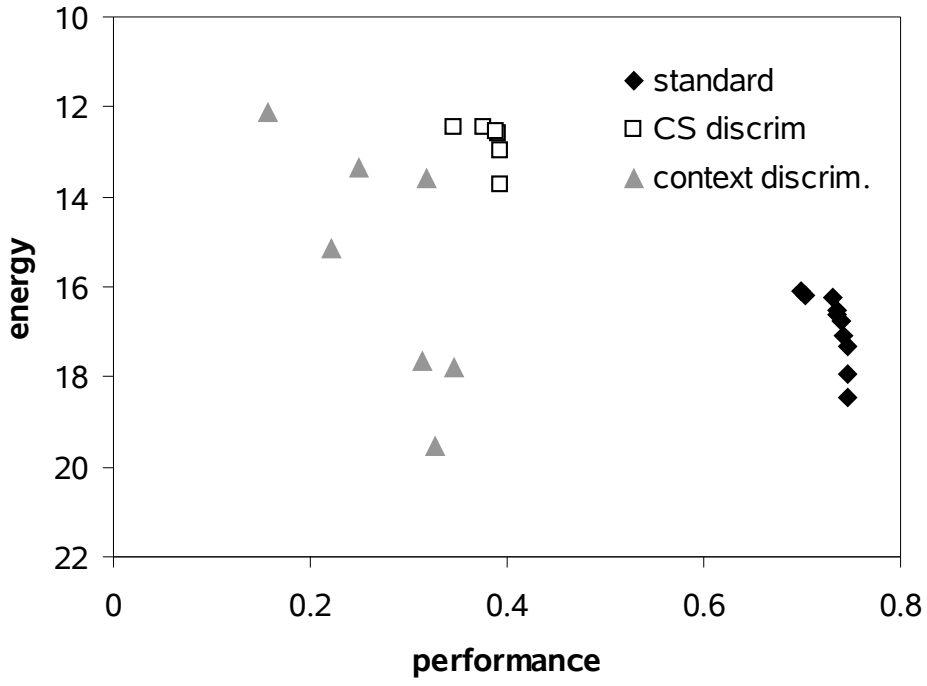

Figure 4: LTP rates on the connections from the cortical, parahippocampal and hippocampal layers to the dorsal striatum, for the three experiments (panels a-c). Gray bars give the minimum value in the population of solutions, white bars give the range of values. Average LTP rates are given by the line inside the white bar.