The Neuroscientist

Information processing in decision-making systems

Journal:	*Neuroscientist*
Manuscript ID:	NRO-11-RE-0048.R2
Manuscript Type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	van der Meer, Matthijs; University of Waterloo, Department of Biology and Centre for Theoretical Neuroscience Kurth-Nelson, Zeb; University College London, Wellcome Trust Centre for Neuroimaging Redish, A.; University of Minnesota, Department of Neuroscience
Keywords:	decision-making, rat, hippocampus, dorsal striatum, ventral striatum
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Supplemental-Movie-R117-2007-06-15-VTE.wmv	

Information processing in decision-making systems

Matthijs van der Meer* Zeb Kurth-Nelson† A. David Redish‡

09 / Dec / 2011

Abstract

Decisions result from an interaction between multiple functional systems acting in parallel to process information in very different ways, each with strengths and weaknesses. In this review, we address three action-selection components of this decision-making system: The *Pavlovian* system releases an action from a limited repertoire of potential actions such as approaching learned stimuli. Like the Pavlovian system, the *habit* system is computationally fast, but permits arbitrary stimulus-action pairings. These associations are a "forward" mechanism; when a situation is recognized, the action is released. In contrast, the *deliberative* system is flexible, but takes time to process. The deliberative system uses knowledge of the causal structure of the world to search forward into the future, planning actions to maximize expected rewards. This depends on the ability to imagine future possibilities, including novel situations, and allows decisions to be taken without having to have previously experienced the options. Various anatomical structures have been identified that carry out the information processing of each of these systems: hippocampus constitutes a map of the world that can be used for searching / imagining the future, dorsal striatal neurons represent situation-action

*mvdm@uwaterloo.ca, University of Waterloo
†zebkurthnelson@gmail.com, Wellcome Trust Centre for Neuroimaging, University College London
‡redish@umn.edu, University of Minnesota

associations, and ventral striatum maintains value representations for all three systems. Each of these systems presents vulnerabilities to pathologies that can manifest as psychiatric disorders. Understanding these systems and their relation to neuroanatomy opens up a deeper way to treat the structural problems underlying various disorders.

Acknowledgements. This work was supported by NIH research grants MH080318 and DA024080 (ADR, ZKN), and by the Canada Research Chairs Program through the National Science and Engineering Research Council [NSERC, Tier II], VENI award 863.10.013 from the Netherlands Organisation for Scientific Research [NWO] and the University of Waterloo (MvdM).

1 Introduction

The brain is an information-processing machine evolved to make decisions: it takes information in, stores it in memory, and uses that knowledge to improve the actions the organism takes. At least three distinct action-selection systems have been identified in the mammalian brain: a Pavlovian action-selection system, a Deliberative action-selection system, and a Habit action-selection system.[1] In this review, we analyze these decision systems from an information processing standpoint. We consider their similarities, differences, and interactions in contributing to a final decision. The Pavlovian action-selection system learns about stimuli that predict motivationally relevant outcomes, such that Pavlovian stimuli come to release actions learned over an evolutionary timescale (Dayan and others, 2006). Although diverse stimuli can participate in Pavlovian learning, the available actions remain limited (e.g. salivate, approach, freeze; Bouton, 2007). Deliberative action-selection is a complex process that includes a search through the expected consequences of possible actions based on a world model. These consequences can then be evaluated online, taking current goals and / or motivational state into account, before selecting an action (Niv and others, 2006). Although Deliberation is very flexible, it is also computationally expensive and slow. The Habit system entails an arbitrary association between a complexly-recognized situation and a complex chain of actions (Sutton and Barto, 1998). Once learned, such cached actions are fast, but can be hard to change.

This review will consist of three parts. In part 1, we will discuss each of the three decision-making systems, with an emphasis on the underlying information-processing steps

[1]Technically, a reflex is also a decision, as it entails the taking of an action in response to stimuli. The fact that a reflex is a decision can be seen in that a reflex only takes the action under certain conditions and it interacts with the other decision-making systems (e.g. it can be overridden by top-down processes,). In the language of decision-making systems developed here, a reflex is a specific action taken in response to a triggering condition. Both the triggering condition and the action taken are learned over evolutionary time-scales. The anatomy, mechanism, and specific stimulus / response pairs associated with reflexes are well understood and available in most primary textbooks and will not be repeated here.

that differentiate them. In part 2, we will discuss specific brain structures and what is known about their individual roles in each of the systems. In part 3, we will discuss some of the implications of the multiple decision-making system theory. Evidence suggests that all three decision-making systems are competing and interacting to produce actions in any given task. We will address the question of how they interact in the discussion in part 3.

Throughout this review, we will concentrate on data from the rat because (1) concentrating on a consistent organism allows better comparisons between systems, and (2) it is the best studied in terms of detailed neural mechanisms for each of the systems. However, these same systems exist in humans and other primates, and we will connect the rat data to primate (human and monkey) homologies when the data is available.

[Box 1 about here.]

2 Action-selection-systems in the mammalian brain

2.1 Pavlovian action-selection

Pavlovian action selection arises because hard-wired, species-specific actions can be governed by associative learning processes (Bouton, 2007). “Unconditioned responses” (URs) classically are physiological responses such as salivation when smelling a lemon or a galvanic skin response following a shock, but also include responses more recognizable as actions, such as approach to a sound, freezing in anticipation of shock, or fleeing from a predator. As organisms learn associative relationships between different events (stimuli including contexts) in the world, originally neutral stimuli (i.e. not capable of evoking an UR) can come to release “conditioned responses”: a bell that predicts food delivery triggers salivation. The bell becomes a conditioned stimulus (CS), to which the organism emits a conditioned response (CR). The action-releasing component of this association depends

on a circuit involving the ventral striatum, the amygdala, and their connections to motor circuits (Ledoux, 2002; Cardinal and others, 2002).

[Figure 1 about here.]

A distinguishing feature of Pavlovian responses is that they occur in the absence of any relationship between the response and subsequent reinforcement. For instance, pigeons typically peck at a cue light predictive of food delivery (CS), even though there is no reward for doing so. Moreover, this so-called "autoshaping" behavior can persist even if the experiment is arranged such that pecking the CS actually reduces reward obtained (Breland and Breland, 1961; Dayan and others, 2006). Thus, Pavlovian actions are selected on the basis of an associative relationship with a particular outcome, rather than on the basis of the action being reinforced.

In rats, Pavlovian action selection is illustrated by comparing *sign-tracking* and *goal-tracking* behavior: when a light signals the availability of food at a separate port, some rats learn to approach the light and chew on it (*sign-tracking*), as if the light itself has gained some food-related concept in the rat's mind. Obviously, the better decision would be to approach the food port when the light turns on (*goal-tracking*). Which rats show sign-tracking and which rats show goal-tracking correlates with and depends on dopamine signals in the ventral striatum (Flagel and others, 2011).

A convenient way of thinking about the mechanism underlying Pavlovian action selection is that the relationship between CS and US gives rise to a neural representation of (aspects of) the US, known as an "expectancy", when the CS is presented; for instance, when the bell is rung, an expectancy of food is produced. To the extent that this expectancy resembles a representation of the US itself, the CR can resemble the UR, but Pavlovian CRs are not restricted to simply replicating the UR. For instance, if the US is devalued (e.g. pairing food with illness) then the CR is strongly attenuated, and different CRs can be produced

depending on the properties of different CSs associated with the same US (Bouton, 2007). Expectancies can have outcome-identity-specific properties (e.g. food vs. water) as well as more general properties (appetitive vs. aversive). These properties interact with current motivational state and the identity of the CS to produce particular CRs.

Furthermore, Pavlovian expectancies can modulate instrumental action selection, an effect termed *Pavlovian-instrumental transfer* (PIT); PIT entails an interaction between motivational components driven by Pavlovian valuation and other action-selection systems (Talmi and others, 2008).

In summary, purely Pavlovian action-selection is characterized by a limited, hard-wired "repertoire" of possible actions, arising from the interplay of an expectancy generated by the CS-US association, motivational state, and actions afforded by the environment (Huys, 2007). Critically, Pavlovian actions can be acquired in the absence of instrumental contingencies, and can therefore be irrelevant or even detrimental to instrumental performance (Breland and Breland, 1961; Dayan and others, 2006). However, expectancies generated through Pavlovian relationships can powerfully modulate instrumental action selection (Talmi and others, 2008).

2.2 Cached-action systems (Habit).

Purely Pavlovian decisions can only release of a limited set of actions. In contrast, the Habit or "cached-action" system forms arbitrary associations between situations and actions, which are learned from experience (Figure 2). Computationally, cached-action system performance entails two deceptively simple steps: recognize the situation and release the associated action. The complexity in cached-action systems arises in the learning process, which must both learn a categorization to recognize situations and also learn which action to take in that situation so as to maximize one's reward.

[Figure 2 about here.]

There are models of both of these components that have been well-integrated with neurophysiology. First, the situation-recognition likely happens through content-addressable mechanisms in cortical systems (Redish and others, 2007). These systems are dependent on the presence of dopamine, particularly for stability of representations. (In the presence of dopamine, situation representations are stable. In the absence of dopamine, situation representations become less stable.)

Second, the association between situation and action is well described by temporal-difference reinforcement learning (TDRL) algorithms (Sutton and Barto, 1998) driven by dopaminergic influences on dorsal (especially dosolateral) striatal systems (Box 2) — the association is trained up by the dopaminergic value-prediction error signal (Schultz and others, 1997). When the value-prediction error is greater than zero, the system should increase its likelihood of taking an action, and when the value-prediction error is less than zero, the system should decrease its likelihood of taking an action. Thus, unlike Pavlovian systems, cached-value system decisions are dependent on a history of reinforcement, that is, they are instrumentally learned. Anatomically, these striatal systems include both *go* (increase likelihood of taking an action) and *no-go* (decrease likelihood of taking an action) systems, each of which are influenced by the presence or absence of a dopaminergic signal (Frank, 2011).

The cached-action system can be seen as a means of shifting the complexity of decision-making from action-selection to situation-recognition. Particularly vivid examples arise in sports. A batter has to decide whether to swing a bat; a quarterback has to decide which receiver to throw to. The action itself is habitual and fast. The hard part is knowing whether this is the right moment to take the action. This arrangement offloads the hard computational work to situation-categorization, which the human brain is extraordinarily good at.

An important prediction of this cached-action learned association is that the dorsolateral association neurons should represent situation-action pairs, but only those pairs that are useful to the animal. From these descriptions, we can make several predictions about these neural representations. (1) They should develop slowly. (2) They should only reflect the current situation. (3) They should only represent information about the world if that information is informative about reward delivery. In the discussion of dorsal striatum, we will see that all three of these predictions are correct descriptions of dorsolateral neural ensembles in the rat. (See Section 3.2.)

The limitations of cached-action systems reside in their inflexibility (Niv and others, 2006). Although a cached-action system can react quickly to a recognized situation, modifying the association takes time and requires extensive experience.[2] Furthermore, the cached-action system is not aware of outcomes (for example, it is insensitive to devaluation); instead, a stimulus or situation leads directly forward into an action without consideration of the consequences. The Deliberative system addresses this limitation.

2.3 Deliberative action-selection

Sometimes, one has to make decisions without having the opportunity to try them out multiple times. Take, for example, a postdoc with two faculty offers, at very different universities in very different locations. That postdoc does not get the opportunity to try each of those two jobs and use any errors in value-prediction to learn the value of each offer. Instead, our intrepid postdoc must imagine himself in each of those two jobs, evaluate the likely rewards and costs associated with those offers, and then make a decision. This is the

[2]There is some evidence that this experience can be achieved without repeating the actual experience through a consolidation process in which the experience is replayed internally (Morris and others, 1982; Redish and Touretzky, 1998; Sutherland and McNaughton, 2000). Computationally, learning through imagined repetition of a specific experience is similar to increasing the learning rate; however, if there is noise in the replayed memories, this can aid in generalization processes in the situation-recognition and association components (Samsonovich and Ascoli, 2005).

process of *deliberation* (Figure 3).

Deliberation requires knowledge of the consequences of one's potential actions: a world model. Computational models have thus termed deliberative processes "model-based" to differentiate them from cached-action processes ("model-free"; Niv and others 2006). Historically, the idea that rats and other animals could deliberate was first proposed by Tolman in the 1930s (Tolman, 1932), but without the available mathematical understanding of information processing, algorithm, or computational complexity, it was impossible to understand how a deliberation system might work. Tolman's hypothesis that rats deliberated over options came from observations originally made by Meunzinger and Gentry in 1931 that under certain conditions, rats would pause at a choice point and turn back and forth, alternately towards the multiple options, before making a decision (Muenzinger and Gentry, 1931). This process was termed "vicarious trial and error" (VTE). VTE events occur after an animal has become familiar with an environment, but when animals are still learning, when they must be flexible about their choices, and when they have to change from a learned habit (Tolman, 1932). Tolman explicitly hypothesized that animals were imagining themselves in the future during "vicarious trial and error".

Imagining oneself in a future is a process called *episodic future thinking* (Atance and O'Neill, 2001; Buckner and Carroll, 2007) and requires an interaction between the hippocampus and the prefrontal cortex (Schacter and Addis, 2011; Hassabis and Maguire, 2011). It entails pulling together concepts from multiple past experiences to create an imagined future (Schacter and Addis, 2011). Because this imagined future is constructed as a coherent whole, only one future tends to be constructed at a time (Atance and O'Neill, 2001). That is, deliberation entails a serial search between options. Also, because this imagined future is constructed, it depends greatly on what aspects of that future event are attended to (Hill, 2008). Attention appears again in the evaluation step because deliberative decisions tend to occur between options with very different advantages and disadvantages. How-

ever, this makes the deliberation process flexible — by changing his attention to compare teaching and research opportunities or to compare lifestyles in the two cities, our postdoc could change the valuation of the two options, before having to make the decision to take one of the two jobs.

As will be discussed below, we now know that during vicarious trial and error (VTE) events, hippocampal representations sweep forward serially through the possibilities (Johnson and Redish, 2007), and both ventral striatal and orbitofrontal reward-related representations covertly signal reward expectations (van der Meer and Redish, 2009, Steiner and Redish, *Society for Neuroscience Abstracts*, 2010). Interestingly, dorsolateral striatal neurons (thought to be involved in cached-action systems) do not show any of these effects (van der Meer and others, 2010).

[Figure 3 about here.]

[Box 2 about here.]

3 Structures involved in decision-making

In a sense, the agent itself is a decision-making machine, and thus the entire brain (and the entire body) is involved in decision-making. However, some of the specific aspects of the action-selection systems detailed above map onto distinct computational roles, mediated by dissociable decision-making circuits in the brain.

3.1 Hippocampus

Tolman suggested that the brain uses a "cognitive map" to support decision-making. In his original conception, this "map" was a representation of both spatial relationships (*if I turn left from here, I will be over there...*), and causal relationships (*if I push this lever, good food*

will appear...) (Tolman, 1932; Johnson and Crowe, 2009). The key to deliberative decision-making is the ability to create a representation of other places and other times (in the case of a coherent, rich representation, this is sometimes called *mental time travel*; Buckner and Carroll 2007; Schacter and Addis 2011). In humans, this ability depends on the hippocampus, whether those other places and other times are in the past (episodic memory, Cohen and Eichenbaum, 1993) or the future (episodic future thinking, Hassabis and Maguire, 2011).

[Figure 4 about here.]

In rats, the primary information encoded by the primary output cells of the hippocampus (excitatory pyramidal cells in CA3 and CA1) is the spatial location of the animal — these are the famous "place" cells (O'Keefe and Nadel, 1978; Redish, 1999, see Figure 4). Hippocampal cells are also sensitive to non-spatial information, but this non-spatial information (such as the presence of a certain object, the color of the walls) modulates the place representation (Redish, 1999).

We will see below that dorsolateral striatal cells encode the information needed to get reward (Section 3.2). Dorsolateral cells do respond to spatial information on spatial tasks, but not on tasks in which the spatial location of the rat is not predictive of reward (Schmitzer-Torbert and Redish, 2008; Berke and Eichenbaum, 2009). In contrast, hippocampal cells show even better spatial representations when the task gets complicated, even when the aspect that makes it complicated is non-spatial (Fenton and others, 2010; Wikenheiser and Redish, 2011).

What are the properties that we expect the hippocampal map to have, in order to be useful for deliberative decision-making? First, a map must be available as soon as possible. Deliberative decision-making is more flexible than habit, and is generally used first when learning new tasks (Killcross and Coutureau, 2003; Redish and others, 2008). Thus, we would expect the hippocampal map to appear quickly, even if it must settle down to stabil-

ity over time. Second, one will need multiple maps for planning in different environments and with different reward distributions. Third, the map should go beyond a simple record of previous experiences; it needs to enable prediction of routes or outcomes that have rarely or not yet been experienced.

In fact, the hippocampal place fields have the appropriate representational firing patterns and the correct dynamics to be the map that is searched during deliberation. The place fields appear from the first entry into an environment (Hill, 1978; Redish, 1999), although they may take time to stabilize, and the stability depends on the need to attend to the task at hand and the presence of dopamine (Kentros and others, 2004). In each environment, there is a random mapping from place cell to place field, such that each cell has a random chance of having a field in an environment and a random location (or locations) of preferred firing in each environment. If the distribution of goal locations within an environment is changed drastically, one sees a dramatic remapping of the place fields. Finally, the distribution of place fields within an environment is approximately uniform: although there is some evidence that place fields are smaller around goals, producing a concentration of place fields around goals (Hollup and others, 2001), place fields do not accumulate around locations that require more complex action selection information (van der Meer and others, 2010).

So, what would deliberation and imagination look like on such a map? During deliberation, animals should pause at choice points and one should see sequential, serial representations of positions sweeping ahead of the animal. These representations should preferentially occur at choice points and preferentially during deliberative rather than habitual events. This is exactly what is seen. As noted above, when rats come to difficult choice points, they pause and look back and forth, showing a behavioral phenomenon called VTE (Muenzinger and Gentry, 1931; Tolman, 1932). During these VTE events, the hippocampal place cells with place fields ahead of the animal fire in sequence, first down one path, then

down the other (Johnson and Redish, 2007). These sequences start at the location of the rat and proceed to the next available goal. (See Figure 4.) These sequences are significantly ahead of the animal, rather than behind it. They are serial, not parallel, and preferentially occur during VTE events at choice points (Johnson and Redish, 2007). These are the neural correlates one would expect from a deliberative search process.

In line with these neural dynamics, hippocampal lesions impair the ability of humans to remember the past (episodic memory, Cohen and Eichenbaum, 1993), to imagine the future (episodic future thinking, Schacter and Addis, 2011), and to plan beyond the present (Hassabis and Maguire, 2011). Hippocampal lesions impair the ability of rats to navigate complex spatial environments (O'Keefe and Nadel, 1978; Redish, 1999) and to place new objects within the environmental context (the schema of the world, Tse and others, 2007), and they attenuate VTE events (Hu and Amsel, 1995). Thus, the hippocampus implements a searchable map comprised of relationships between spatial locations, objects, and contexts. The dynamics of hippocampal representations match the expectations one would see if the hippocampus was involved in planning.

3.2 Dorsal Striatum

The functional roles of the striatal subregions reflect the topographical organization of its inputs and outputs (Swanson, 2000). A distinction is generally made between the dorso-lateral striatum, interconnected with sensory and motor cortex, the dorsomedial striatum, interconnected with associative cortical areas, and the ventral striatum, interconnected with hippocampal and frontal cortical areas; this subdivision should be understood as gradual, rather than as clear-cut and abrupt. In addition, this connectivity-based subdivision should also be understood as having additional effects along the anterior-posterior axis (Swanson, 2000; Yin and Knowlton, 2004).

Lesion studies indicate a striking dissociation between dorsolateral and dorsomedial

striatum, with dorsolateral striatum being important for the performance of habitual actions, and dorsomedial striatum being important for the performance of deliberative (goal-directed, going to a place rather than making a response) actions (Yin and Knowlton, 2004). Strikingly, Atallah and others (2007) found that dorsolateral striatum was required for the performance, but not acquisition, of an instrumental S-A (habit) task.

Recording studies have tended to concentrate on the anterior dorsolateral striatum because lesion studies have found that the anterior dorsolateral striatum produced contrasting effects to hippocampal lesions on tasks that put place-directed (deliberative) strategies in conflict with response-directed (habit) strategies. See Figure 7. Studies in the anterior dorsolateral striatum find that cells learn to encode situation-action pairs, such that the situation-correlations depend on the information necessary to find reward (Berke and Eichenbaum, 2009; Schmitzer-Torbert and Redish, 2008). Cells in the anterior dorsolateral striatum develop task-related firing with experience (Barnes and others, 2005; van der Meer and others, 2010). This task-related firing tends to occur at task-components where habitual decisions needed to be initiated (Barnes and others, 2005). There has not been much equivalent recording done in posterior dorsomedial striatum, although one recent study did find that in anterior dorsomedial striatum, cells developed task-related firing more quickly than anterior dorsolateral striatum and that these cells showed firing related to decisions (Thorn and others, 2010).

[Figure 5 about here.]

An influential model of the learning and performance of habitual actions in the dorsolateral striatum is that it provides situation-action associations in a *model-free* temporal difference reinforcement learning algorithm (TDRL, see Box 2). This conceptualization suggests that the dorsal striatum associates situation information coming from cortical structures with actions as trained by the dopaminergic training signals.

Recording studies in anterior dorsolateral striatum have identified striatal activity related to the internal variables needed for the action-selection component of TDRL in both rat (Schmitzer-Torbert and Redish, 2008; Barnes and others, 2005; Berke and Eichenbaum, 2009; van der Meer and others, 2010) and monkey (Hikosaka and others, 1989; Samejima and others, 2005; Lau and Glimcher, 2007). A particularly striking effect, observed in different tasks, is the emergence of elevated dorsolateral striatal activity at the beginning and end of action sequences (Barnes and others, 2005; Thorn and others, 2010) and the separation of action-related and reward-related activity in anterior dorsolateral striatum (Schmitzer-Torbert and Redish, 2004, 2008). These results suggest network reorganization with repeated experience consistent with the development of habitual behavior. Comparisons of dorsolateral and dorsomedial striatal activity have yielded mixed results (Kimchi and Laubach, 2009; Thorn and others, 2010; Stalnaker and others, 2010), but generally, these studies have compared anterior dorsolateral striatum with anterior dorsomedial striatum. It is not clear that these studies have directly tested the differences in information processing in different dorsal striatal components under deliberative and habit-based decision-making.

3.3 Ventral Striatum

The ventral aspect of the striatum (encompassing the core and shell of the nucleus accumbens, the ventral caudate / putamen, and the olfactory tubercule) is a heterogenous area anatomically defined through its interconnections with a number of “limbic” areas (Swanson, 2000). Historically, ventral striatum has long been seen as the gateway from limbic structures to action components (Mogenson and others, 1980). Critically, ventral striatum is a major input to dopaminergic neurons in the ventral tegmental area (VTA) which in turn furnishes ventral striatum itself, dorsal striatal areas, prefrontal cortex, and the hippocampus with dopamine signals. The close association with dopamine and convergence of limbic inputs renders vStr a central node in brain networks processing reward- and motivation-

related information.

[Figure 6 about here.]

The close anatomical and functional association with dopamine (both in terms of providing input to the VTA, but also because of its dense return projection; Haber 2009) means that major views of ventral striatum function are intertwined with dopamine function. One such idea is that the ventral striatum computes the value of situations (which includes rewards actually received as well as discounted future rewards expected; see Box 2) to supply one term of the prediction error equation to the VTA. The VTA prediction error, in turn, serves to update ventral striatal representations of values of given situations as well as dorsal striatal representations of the values of taking actions. This casts the role of ventral striatum as supporting gradual learning from feedback, as is thought to occur in the "habit" system; experimental support for this notion comes, for instance, from inactivation studies that find large effects on acquisition, but small effects (if any) on performance (Atallah and others, 2007). However, recent demonstrations that the dopamine input to the ventral striatum is not homogenous (Aragona and others, 2009) pose a challenge for a straightforward mapping onto TDRL's conception of the error signal as a single value.

Ventral striatum is also importantly involved in the more immediate modulation of behavior — it mediates aspects of Pavlovian conditioned responding, including autoshaping (Cardinal and others, 2002) and sensitivity to devaluation of the US (Singh and others, 2010). It is also required for conditioned reinforcement (willingness to work to receive a CS, Cardinal and others, 2002). Value representations in ventral striatum are also important for the deliberative and habit systems. Two neuronal firing correlates have been reliably found in ventral striatum: reward-related firing that occurs shortly after an animal receives reward (Lavoie and Mizumori, 1994; Taha and Fields, 2005; van der Meer and Redish, 2009) and "ramp" neurons that increase firing as an animal approaches a reward (Lavoie and Mizu-

mori, 1994; van der Meer and Redish, 2011).

For representations of potential future states – as required by deliberation and found in the hippocampus – to be useful in deliberative decision-making, some kind of evaluation of the value of these imagined states is required. One possibility for such evaluation is that future states (represented in the hippocampus) function as a cue or state input for the on-line computation of ventral striatal values. Consistent with this possibility, recording studies in ventral striatum show close association with hippocampal inputs, including re-activation of reward neurons in sync with replay of hippocampal activity during sleep and rest (Lansink and others, 2009), and re-activation of reward-related firing on movement initiation and during deliberative decision-making (van der Meer and Redish, 2009). Intriguingly, there is evidence that dopamine inputs to ventral striatum are particularly important for the performance of "flexible" approach behavior likely to require such an on-line evaluation process, but not for the performance of a similar, but stereotyped, version of the task (Nicola, 2010).

In order to train the habit-based situation-action association, one would also need a value signal, such as that provided by the TDRL value-learning system. Ventral striatal ramp cells show the right firing patterns to provide this signal. (See Figure 6.) Although ventral striatum is often necessary for learning, it is not necessary for performance of habit-based instrumental decision tasks (Atallah and others, 2007).

In sum, value plays a central role in Pavlovian, habit, and deliberative systems alike, and as a central node in reward processing, it appears that ventral striatum plays a role in all three systems. To what extent this role reflects unitary processing (the same computational role) or different processing for each system, and how this relates to known heterogeneities such as core / shell, are important, unanswered, current research topics.

[Figure 7 about here.]

4 Discussion

4.1 Decision-making and neuroeconomics

Neuroeconomics attempts to study decision-making starting from the point of view of microeconomics, relating neuroscientific results to economic variables. The neuroeconomic view of decisions is that each available outcome is evaluated to a scalar "value" or "utility", and these scalars are compared, with a preference for choosing higher-value outcomes.

The multiple systems theory postulates that each system has its own decision-making algorithms, which compete and interact to produce the actual decision. This seems to be at odds with the neuroeconomic view that there is a unitary evaluation of each outcome. One can imagine at least two different ways of reconciling these views. Perhaps neuroeconomic valuation is a descriptive approximation for the overall behavior that emerges from multiple systems interacting. Or, perhaps neuroeconomic valuation is used within some of the multiple decision-making systems, but can be violated when other systems take over (Figure 8). We suggest that the latter is the case.

[Figure 8 about here.]

Many of the experiments identifying neural correlates of value use habitual tasks and carefully eliminate Pavlovian influences. However, Pavlovian influences can undermine neuroeconomic valuation. For example, real options (such as a physical candy bar) are harder to reject than linguistically labeled options (Boysen and Berntson, 1995; Bushong and others, 2010) — that is, it is easier to say "I will keep my diet and not eat that candy bar" when the candy bar is not in front of you. Similarly, pigeons cannot learn not to peck in order to get reward (Breland and Breland, 1961), and chickens cannot learn to run away to get food (Hershberger, 1986). The unified neuroeconomic account would indicate that, once the animals have learned the task contingencies, they should make the action that leads to

the larger reward. Thus, it would follow from a neuroeconomic standpoint that the animals are simply unable to learn the task contingency, a possibility made less likely by the fact that, in the Boysen and Berntson (1995) experiment, the same chimpanzees could learn to point to an Arabic numeral to receive the larger pile of candy. The multiple systems theory provides the more satisfying account: that animals do learn the task, but when a food reward is within pointing distance, a Pavlovian unconditioned response (reaching / pointing) is released, which wins out over a more rational choice in the competition between systems.

Of course, the organism remains a unitary being — eventually there must be a decision made at the level of the muscles. An interesting (and as yet unanswered) question is whether the integration of the multiple decision-making systems happens at the level of the brain, before action-commands are sent down to the spinal cord, or whether the final integration only happens at the level of the motor commands themselves. (Most likely, some integration happens at each stage.)

4.2 Computational psychiatry

The ability to identify specific mechanisms of decision-making provides a potential mechanistic language to address how decisions can go wrong (Redish and others, 2008; Maia and Frank, 2011; Huys and others, 2011). Psychiatry has historically been based on categorizations of observable symptoms, which may or may not have direct relevance to underlying mechanistic causes (McHugh and Slavney, 1998). The multiple decision-making systems theory provides a level of structure to connect information processing mechanisms in the brain with observable behavior. Now that we can talk about specific mechanisms, it becomes possible in this mechanistic language to describe various things that can go wrong.

Of course, this only works as long as the description still maps on to the way things are functioning. For example, the pathology could be massive brain trauma or neurodegeneration to such an extent that “Pavlovian decision-making” is no longer a meaningful descrip-

tion of the biological system. But we suggest that many psychiatric disorders, including autism, borderline personality disorder (BPD), depression, and addiction are meaningfully described as parameter variations within multiple decision-making systems (Huys, 2007; Redish and others, 2008; Kishida and others, 2010).

In classic psychiatry, disease states are clustered by their distance in a symptom space which arose historically by phenomenological description (McHugh and Slavney, 1998). Additionally, the same binary diagnosis can be given for very different symptom combinations, because diagnoses are made when any n of m symptoms are present. For example, there are nine criteria in the DSM-IV for borderline personality disorder (BPD), and there is a positive diagnosis when five or more of these criteria are met. Thus two people could have only one criterion in common and receive the same diagnosis of BPD (and likely be offered the same pharmacological treatment, when the underlying anatomical and neuromodulatory pathologies may be completely different). The superficiality of the symptom space is analogous to diagnosing "chest pain". A deeper understanding of mechanism reveals that either acid reflux or heart disease can cause chest pain. Likewise, computational psychiatry argues that psychiatric disorders ought to be classified based on their distance in "causal" or "functional" space, and treated based on an understanding of the links between the anatomy and physiology of the brain and the dimensions of this mechanistic space.

In each decision-making system in the brain, there are parameters which, when set inappropriately, produce maladaptive decisions — in other words, vulnerabilities (Redish and others, 2008). Drug-addiction, for example, has been partially modeled as a disorder of the habit (cached-action) system (Redish, 2004). We saw that the phasic firing of dopaminergic neurons encodes the reward prediction error signal of TDRL (Schultz and others, 1997). Since many drugs of abuse share the common mechanism of boosting phasic dopamine firing to mediate their reinforcing effects, it is logical that these drugs are pharmacologically manipulating the computations in the learning process to produce an uncompensable

prediction error, such that the reward expectation following drug-seeking actions is perpetually revised upwards. However, there are features of addiction that extend beyond habit. Addicts will sometimes engage in complex planning (deliberation) to obtain drugs (Heyman, 2009). There are differences in how important the dopamine signal is to different users' taking of different drugs (Badiani and others, 2011). This suggests that addiction is also accessing vulnerabilities in the deliberative and other systems (Redish and others, 2008).

Similarly, depression has also been suggested to have roots in deliberative decision-making processes (Huys, 2007). In deliberative decision-making, the agent attempts to make inferences about the future consequences of its actions. A key feature of depression is the sense of "helplessness": a belief that the agent has little control over the future reinforcers it will receive. Thus, if we make the assumption of normative inference, we can predict the types of prior beliefs (perhaps genetically modulated) or the kinds of experiences that would lead an agent into periods of depression (Huys, 2007).

Although we have not discussed the "support structures" of motivation and situation-categorization here, both systems can have their own failure modes, which can drive decision-making errors (Flagel and others, 2011; Redish and others, 2007). For example, both the cached-action and deliberative systems require some form of dimensionality reduction of the input space in order to learn situation-action mappings (cached-action), or to search over (deliberative). Pathologies in this cognitive state classification system can also be described computationally. It has been proposed that in problem gambling, agents classify wins as consequences of their actions but attribute losses to ancillary factors (Langer and Roth, 1975).

With an understanding of psychiatric conditions at this mechanistic level, we can start to make more reasoned predictions about what kinds of treatment will be most effective for each individual. The multiple decision-making theory takes us one step closer to that

mechanistic level.

References

Addis DR, Wong AT, Schacter DL (2007) Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45(7):1363–1377.

Aragona BJ, Day JJ, Roitman MF, Cleveland NA, Wightman RM, Carelli RM (2009) Regional specificity in the real-time development of phasic dopamine transmission patterns during acquisition of a cue-cocaine association in rats. European Journal of Neuroscience pp. 1–11.

Atallah EH, Lopeq-Paniagua D, Rudy JW, O'Reilly RC (2007) Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nature Neuroscience 10(1):126–131.

Atance CM, O'Neill DK (2001) Episodic future thinking. Trends in Cognitive Sciences 5(12):533–539.

Badiani A, Belin D, Epstein D, Calu D, Shaham Y (2011) Opiate versus psychostimulant addiction: the differences do matter. Nature Reviews Neuroscience 12(11):685–700.

Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature 437:1158–1161.

Bellman R (1958) On a routing problem. Quarterly Journal of Applied Mathematics 16(1):87–90.

Berke JD, Eichenbaum H (2009) Striatal versus hippocampal representations during win-stay maze performance. Journal of Neurophysiology 101(3):1575–1587.

Bouton ME (2007) Learning and Behavior: a contemporary synthesis. Sinauer Associates.

Boysen ST, Berntson GG (1995) Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (pan troglodytes). Journal of Experimental Psychology: Animal Behavior Processes 21(1):82–86.

Breland K, Breland M (1961) The misbehavior of organisms. American Psychologist 16(11):682–684.

Buckner RL, Carroll DC (2007) Self-projection and the brain. Trends in Cognitive Sciences 11(2):49–57.

Bushong B, King LM, Camerer CF, Rangel A (2010) Pavlovian processes in consumer choice: The physical presence of a good increases willingness-to-pay. American Economic Review 100(4):1556–1571.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience & Biobehavioral Reviews 26(3):321–352.

Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. Annual Reviews Neuroscience 33:269–298.

Cohen NJ, Eichenbaum H (1993) Memory, Amnesia, and the Hippocampal System. MIT Press.

Craig AD (2003) Interoception: the sense of the physiological condition of the body. Current Opinion in Neurobiology 13(4):500–505.

D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319(5867):1264–1267.

Dayan P, Niv Y (2008) Reinforcement learning: The good, the bad and the ugly. Current Opinion in Neurobiology 18(2):185–196.

Dayan P, Niv Y, Seymour B, Daw ND (2006) The misbehavior of value and the discipline of the will. Neural Networks 19:1153–1160.

Fenton AA, Lytton WW, Barry JM, Lenck-Santini PP, Zinyuk LE, Kubik S, Bures J, Poucet B, Muller RU, Olypher AV (2010) Attention-like modulation of hippocampus place cell discharge. Journal of Neuroscience 30(13):4613–4625.

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PEM, Akil H (2011) A selective role for dopamine in stimulus-reward learning. Nature 469(7328):53–57.

Frank MJ (2011) Computational models of motivated action selection in corticostriatal circuits. Current Opinion in Neurobiology 21(3):381–386.

Gupta AS (2011) Behavioral Correlates of Hippocampal Neural Sequences. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Haber SN (2009) Anatomy and connectivity of the reward circuit. In: Handbook of Reward and Decision Making (Dreher JC, Tremblay L, eds.), pp. 3–27. Academic Press.

Hassabis D, Maguire EA (2011) The contruction system in the brain. In: Predictions in the brain: using our past to generate a future (Bar M, ed.), pp. 70–82. Oxford University Press.

Hershberger WA (1986) An approach through the looking glass. Learning and Behavior 14(4):443–451.

Heyman G (2009) Addiction: A disorder of choice. Harvard.

Hikosaka O, Sakamoto M, Usui S (1989) Functional properties of monkey caudate neurons. Journal of Neurophysiology 61(4):780–832.

Hill AJ (1978) First occurrence of hippocampal spatial firing in a new environment. Experimental Neurology 62:282–297.

Hill C (2008) The rationality of preference construction (and the irrationality of rational choice). Minnesota Journal of Law, Science, and Technology 9(2):689–742.

Hollup SA, Molden S, Donnett JG, Moser MB, Moser EI (2001) Accumulation of hippocampal place fields at the goal location in an annular watermaze task. Journal of Neuroscience 21(5):1635–1644.

Hu D, Amsel A (1995) A simple test of the vicarious trial-and-error hypothesis of hippocampal function. PNAS 92:5506–5509.

Huys QJM (2007) Reinforcers and Control: Towards a Computational Ætiology of Depression. Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London.

Huys QJM, Moutoussis M, Williams J (2011) Are computational models of any use to psychiatry. Neural Networks 24(6):544–551.

Johnson A, Crowe DA (2009) Revisiting Tolman, his theories and cognitive maps. Cognitive Critique 1:43–72.

Johnson A, Jackson J, Redish AD (2008) Measuring distributed properties of neural representations beyond the decoding of local variables — implications for cognition. In: Mechanisms of information processing in the Brain: Encoding of information in neural populations and networks (Hölscher C, Munk MHJ, eds.), pp. 95–119. Cambridge University Press.

Johnson A, Redish AD (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. Journal of Neuroscience 27(45):12176–12189.

Kentros CG, Agnihotri NT, Streater S, Hawkins RD, Kandel ER (2004) Increased attention to spatial context increases both place field stability and spatial memory. Neuron 42:283–295.

Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontral cortex of rats. Cerebral Cortex 13(8):400–408.

Kimchi EY, Laubach M (2009) Dynamic Encoding of Action Selection by the Medial Striatum. Journal of Neuroscience 29(10):3148–3159.

Kishida KT, King-Casas B, Montague PR (2010) Neuroeconomic approaches to mental disorders. Neuron 67(4):543–554.

Kurth-Nelson Z, Redish AD (in press) Modeling decision-making systems in addiction. In: Computational Neuroscience of Drug Addiction (Gutkin B, Ahmed SH, eds.). Springer.

Langer EJ, Roth J (1975) Heads i win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. Journal of Personality and Social Psychology 32(6):951–955.

Lansink CS, Goltstein PM, Lankelma JV, McNaughton BL, Pennartz CMA (2009) Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biology 7(8):e1000173.

Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. Journal of Neuroscience 27(52):14502–14514.

Lavoie AM, Mizumori SJY (1994) Spatial-, movement- and reward-sensitive discharge by medial ventral striatum neurons in rats. Brain Research 638:157–168.

Ledoux J (2002) The Synaptic Self. Penguin.

Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. Nature Neuroscience 14(2):154–1652.

McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. Journal of Neuroscience 31(7):2700–2705.

McHugh PR, Slavney PR (1998) The Perspectives of Psychiatry. Johns Hopkins.

Mogenson GJ, Jones DL, Yim CY (1980) From motivation to action: Functional interface between the limbic system and the motor system. Progress in Neurobiology 14:69–97.

Morris RGM, Garrud P, Rawlins JNP, O'Keefe J (1982) Place navigation impaired in rats with hippocampal lesions. Nature 297:681–683.

Muenzinger KF, Gentry E (1931) Tone discrimination in white rats. Journal of Comparative Psychology 12(2):195–206.

Nicola SM (2010) The flexible approach hypothesis: Unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. Journal of Neuroscience 30(49):16585–16600.

Niv Y, Joel D, Dayan P (2006) A normative perspective on motivation. Trends in Cognitive Sciences 10(8):375–381.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304(5669):452–454.

O'Keefe J, Nadel L (1978) The Hippocampus as a Cognitive Map. Oxford: Clarendon Press.

Redish AD (1999) Beyond the Cognitive Map: From Place Cells to Episodic Memory. Cambridge MA: MIT Press.

Redish AD (2004) Addiction as a computational process gone awry. Science 306(5703):1944–1947.

Redish AD, Jensen S, Johnson A (2008) A unified framework for addiction: vulnerabilities in the decision process. Behavioral and Brain Sciences 31:415–487.

Redish AD, Jensen S, Johnson A, Kurth-Nelson Z (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. Psychological Review 114(3):784–805.

Redish AD, Touretzky DS (1998) The role of the hippocampus in solving the Morris water maze. Neural Computation 10(1):73–111.

Rich EL, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. Journal of Neuroscience 29(22):7208–7219.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310(5752):1337–1340.

Samsonovich AV, Ascoli GA (2005) A simple neural network model of the hippocampus suggesting its pathfinding role in episodic memory retrieval. Learn Mem 12(2):193–208.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. Science 300(5626):1755–1758.

Schacter DL, Addis DR (2011) On the nature of medial temporal lobe contributions to the constructive simulation of future events. In: Predictions in the brain: using our past to generate a future (Bar M, ed.), pp. 58–69. Oxford University Press.

Schmitzer-Torbert NC, Redish AD (2004) Neuronal activity in the rodent dorsal striatum in sequential navigation: Separation of spatial and reward responses on the multiple-T task. Journal of Neurophysiology 91(5):2259–2272.

Schmitzer-Torbert NC, Redish AD (2008) Task-dependent encoding of space and events by striatal neurons is dependent on neural subtype. Neuroscience 153(2):349–360.

Schultz W, Dayan P, Montague R (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Singh T, McDannald MA, Haney RZ, Cerri DH, Schoenbaum G (2010) Nucleus accumbens core and shell are necessary for reinforcer devaluation effects on Pavlovian conditioned responding. Frontiers in Integrative Neuroscience 4:126.

Stalnaker T, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2010) Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. Frontiers in Integrative Neuroscience 4(12).

Sutherland GR, McNaughton BL (2000) Memory trace reactivation in hippocampal and neocortical neuronal ensembles. Current Opinion in Neurobiology 10(2):180–6.

Sutton RS, Barto AG (1998) Reinforcement Learning: An introduction. Cambridge MA: MIT Press.

Swanson LW (2000) Cerebral hemisphere regulation of motivated behavior. Brain Research 886(1-2):113–164.

Taha SA, Fields HL (2005) Encoding of palatability and appetitive behaviors by distinct neuronal populations in the nucleus accumbens. Journal of Neuroscience 25(5):1193–1202.

Talmi D, Seymour B, Dayan P, Dolan RJ (2008) Human pavlovian-instrumental transfer. J Neurosci 28(2):360–368.

Thorn CA, Atallah H, Howe M, Graybiel AM (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. Neuron 66(5):781–795.

Tolman EC (1932) Purposive Behavior in Animals and Men. New York: Appleton-Century-Crofts.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM (2007) Schemas and memory consolidation. Science 316(5821):76–82.

van der Meer MAA, Johnson A, Schmitzer-Torbert NC, Redish AD (2010) Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. Neuron 67(1):25–32.

van der Meer MAA, Redish AD (2009) Covert expectation-of-reward in rat ventral striatum at decision points. Frontiers in Integrative Neuroscience 3(1):1–15.

van der Meer MAA, Redish AD (2011) Theta phase precession in rat ventral striatum links place and reward information. Journal of Neuroscience 31(8):2843–2854.

Wikenheiser AM, Redish AD (2011) Changes in reward contingency modulate the trial-to-trial variability of hippocampal place cells. Journal of Neurophysiology 106(2):589–598.

Yin HH, Knowlton BJ (2004) Contributions of striatal subregions to place and response learning. Learning and Memory 11(4):459–463.

Box 1. Functional subsystems. The concept of multiple functional systems should not be taken to imply that there are truly separable "modules" — these systems depend on interactions among multiple brain structures, each of which is providing a different computational component. A useful analogy is that of a hybrid gas / electric car: although there are two separate systems, which depend on dissociable components, both systems also share many components. The car, for example, has only one drive train. Similarly, the car requires numerous other support systems that are shared between the two components, such as the steering system. We would therefore predict that while there will be dissociations in both the information processing and effects of lesions between the systems (van der Meer and others, 2010; Yin and Knowlton, 2004), individual anatomy structures will also be shared between the systems, although they may provide different computational components to each system. For example, the ventral striatum seems to be involved in all three components, including providing mechanisms to re-evaluate changes in Pavlovian value (McDannald and others, 2011), covert representations of valuation during deliberative events (van der Meer and Redish, 2009), and training up Habit systems (Atallah and others, 2007).

There are five functional subsystems that can be indentified as playing roles in decision-making: Pavlovian action-selection, Habit-based action-selection, Deliberative action-selection, and the Motivational and Situation-Recognition support systems. Although our knowledge of the anatomical instantiations of these systems is still obviously incomplete and the roles played by each structure in each functional subsystem are still an area of active research, we can make some statements about components known to be important in each subsystem.

The **Pavlovian** system (pink, Figure 1) includes the periaqueductal gray (PAG), the ventral tegmental area (VTA), the amygdala (AMG), the ventral striatum (vStr), and

the orbitofrontal cortex (OFC) (Ledoux, 2002; McDannald and others, 2011). The **Habit** system (orange, Figure 2) includes the substantia nigra pars compacta (SNc), the dorsolateral striatum (dlStr), the ventral striatum (vStr), and likely the motor cortex (MC) (Yin and Knowlton, 2004; Cisek and Kalaska, 2010). The **Deliberative** system (blue, Figure 3) includes the hippocampus (HC), the prefrontal cortex (PFC), the ventral striatum (vStr), and likely the ventral tegmental area (VTA), and the dorsomedial striatum (dmStr) (Johnson and Redish, 2007; van der Meer and Redish, 2009; Schacter and Addis, 2011; Yin and Knowlton, 2004). In addition, decision-making involves several support structures, not discussed in depth in this review, a **Motivation** system, likely including the hypothalamus (HyTM), the ventral striatum, and the insula and cortical visceral areas (Craig, 2003; Sanfey and others, 2003), and a **Situation categorization** system, likely including most of neocortex (Redish and others, 2007).

Box 2. Temporal difference Reinforcement Learning (TDRL) in three systems. Current theories of reinforcement learning are based on the concept of the *temporal difference* rule. The basic concept of this system is that an *agent* (a person, animal, or computer simulation) traverses a *state-space* of situations. In many simulations, this state-space is provided to the simulation, but real agents (animals or humans) need to determine the situations and their relationships. (*What is the important cue in the room you are in right now?*) Differences in the interpretation of that state-space can produce dramatic differences in decision-making (Kurth-Nelson and Redish, in press). Different forms of TDRL have been applied to each of the decision-making systems.

Pavlovian. "Blocking" experiments demonstrated that if animal has learned that CS1 predicts a certain US, then pairing CS1+CS2 with the US does not result in a CR to CS2 subsequently presented alone (Bouton, 2007). (However, if aspects of the US change, then the second CS will gain associations related to the observed changes, Bouton, 2007; McDannald and others, 2011.) Rescorla and Wagner (1972) proposed that Pavlovian learning requires a *prediction error*: a mismatch between what is expected and what occurs. Since in the blocking experiment, the US is fully predicted by CS1, no CS2-US association develops. In the 1990s, Sutton and Barto showed that this is a special case of the temporal difference learning rule, in which one associates value with situations through a *value-prediction-error signal* (Sutton and Barto, 1998). The temporal-difference rule maintains an estimated future reward value for each recognized situation, such that prediction errors can be computed for any transition between situations, not just for those resulting in reward. Neurophysiological recordings of the firing of dopamine neurons and fMRI BOLD signals of dopamine-projection areas have been shown to track the value-prediction-error signal in Pavlovian conditions remarkably accurately

(D'Ardenne and others, 2008). Flagel and colleagues have found that dopamine release in the core of the nucleus accumbens (the ventral striatum) of sign-tracking rats (but not goal-tracking rats) matches this value-prediction-error signal and that only sign-tracking rats (not goal-tracking rats) can use the CS as a subsequent CR for secondary conditioning (Flagel and others, 2011).

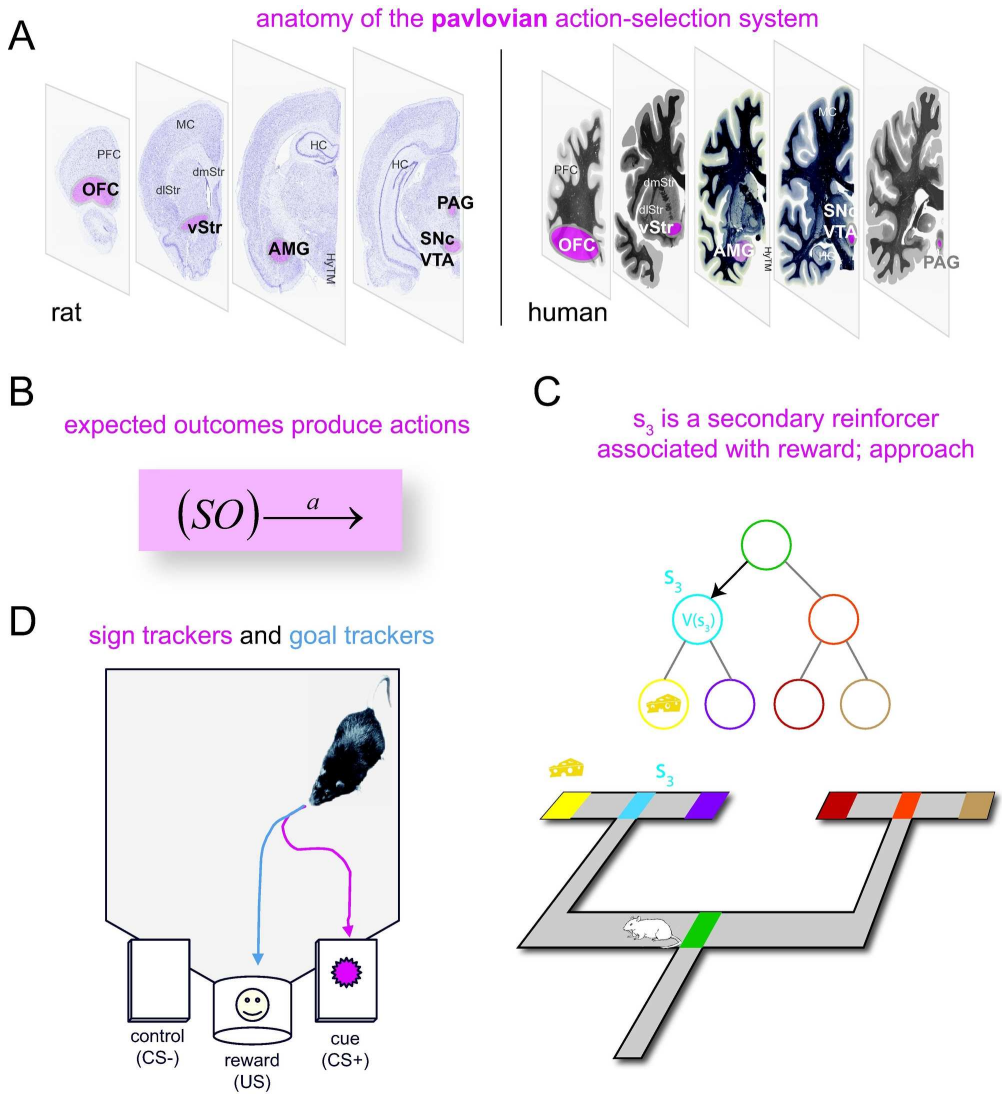
Habit. In the TDRL literature, "habit" learning corresponds to the original temporal difference rule originally proposed by Bellman in 1958 and introduced into the literature by Sutton and Barto (Bellman, 1958; Sutton and Barto, 1998). In the most likely formulation (known as the *actor-critic architecture*), one component learns to predict the value of actions taken in certain situations based on differences between observed value and expected value. That difference signal is also used to train up situation-action associations. It can be shown that under the right conditions of exploration and stationarity, this architecture will converge (eventually) on the decision-policy that maximizes the total reward available in the task (Sutton and Barto, 1998); however, this can take many trials and is inflexible in non-stationary worlds (Dayan and Niv, 2008).

Deliberative. In the TDRL literature, "deliberative" decision-making is based on the concept of *model-based* TDRL (Sutton and Barto, 1998). Here, the agent is assumed to have a model of the causal structure of the world, which it can use to predict the consequences of its actions. From these predictions, the agent can evaluate those expected consequences at the time of the decision, taking into account its current needs and desires (Niv and others, 2006). This hypothesis predicts that deliberative decision-making will be slow (because it requires search, prediction, and evaluation steps; van der Meer and others 2010), and that representations of hypothesized outcomes and covert representations of reward expectation will be detectable in structures critical for deliberative

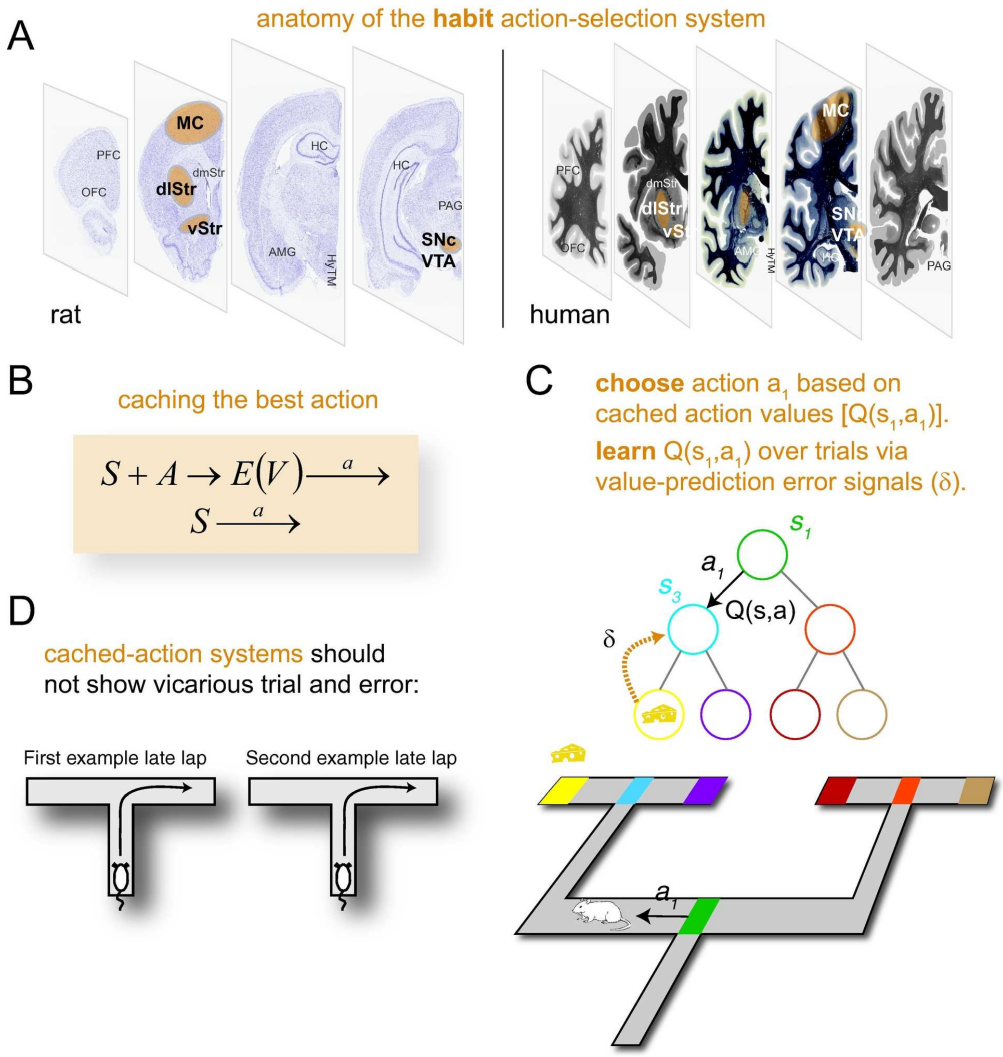
decision-making. As is discussed in the main text, such predictive and covert representations have been found in the hippocampus, ventral striatum, and orbitofrontal cortex (Johnson and Redish, 2007; van der Meer and Redish, 2009; van der Meer and others, 2010, Steiner and Redish, *Society for Neuroscience Abstracts*, 2010).

List of Figures

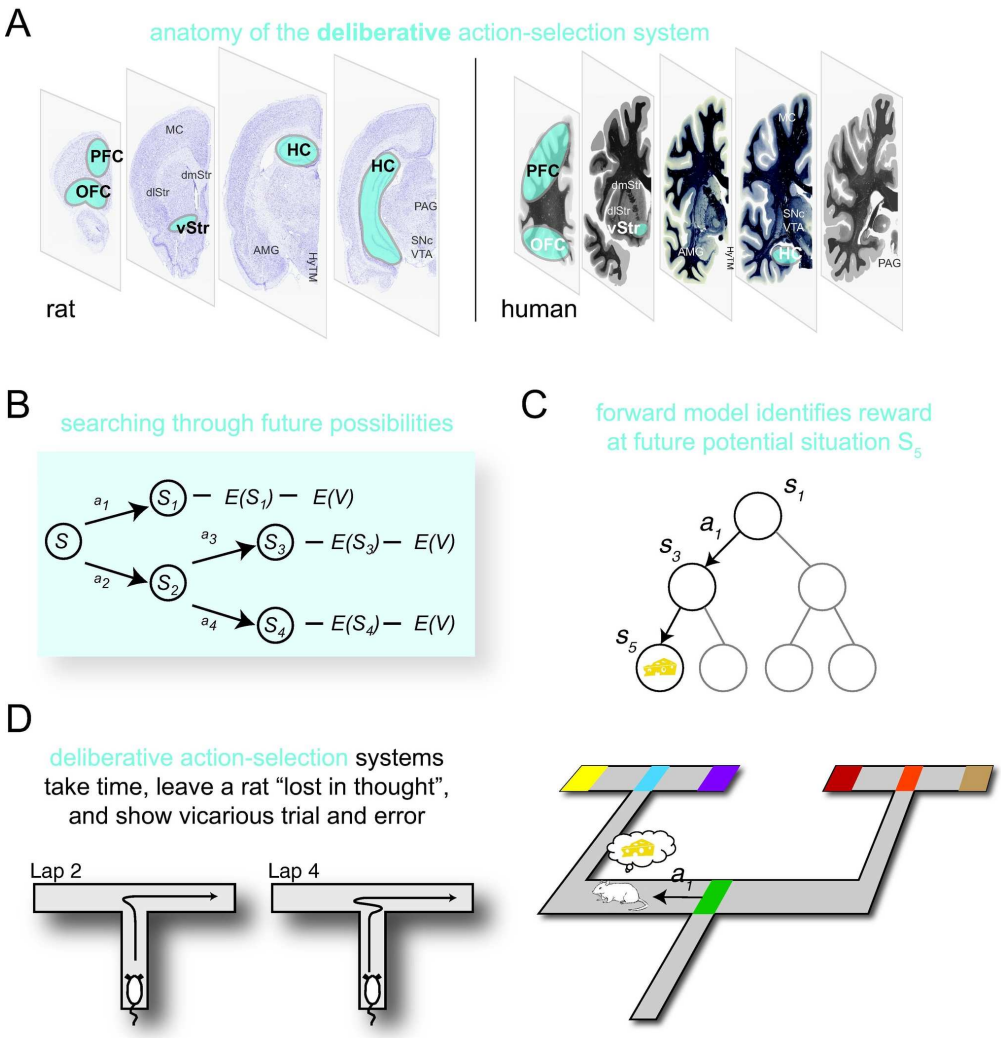
4 **Hippocampal contributions to decision-making.** (A) The hippocampus encodes a map of the environment through the activity across the place cells. (B) Two example place cells from a choice-task. The animal runs north through the central stem, turns right or left at the top of the maze, and receives food on the right or left return rails depending on a complex decision-making criterion. (C) The existence of this map allows imagination and planning through the firing of cells with place fields away from the animal. (D) An example planning sequence. The top panel shows the same maze as in panel (B), with each spike from each cell that fires within a single 150 ms theta cycle plotted at the center of that cell's place field. Colors indicate time in the single theta cycle. The bottom panel shows the firing of the same cells, ordered by their place fields around the maze, with the theta cycle in the local field potential beneath. (E) Both remembering the past and imagining the future activate hippocampus in humans. Subjects were instructed to initially imagine or remember an event (construction) and then to bring to mind as many details about that event as possible (elaboration). Compared to a control task, hippocampus was differentially active during both of these processes. Data in panels (B) and (D) from Gupta (2011), used with permission. Data in panel (E) from Addis and others (2007), used with permission of author and publisher. 45



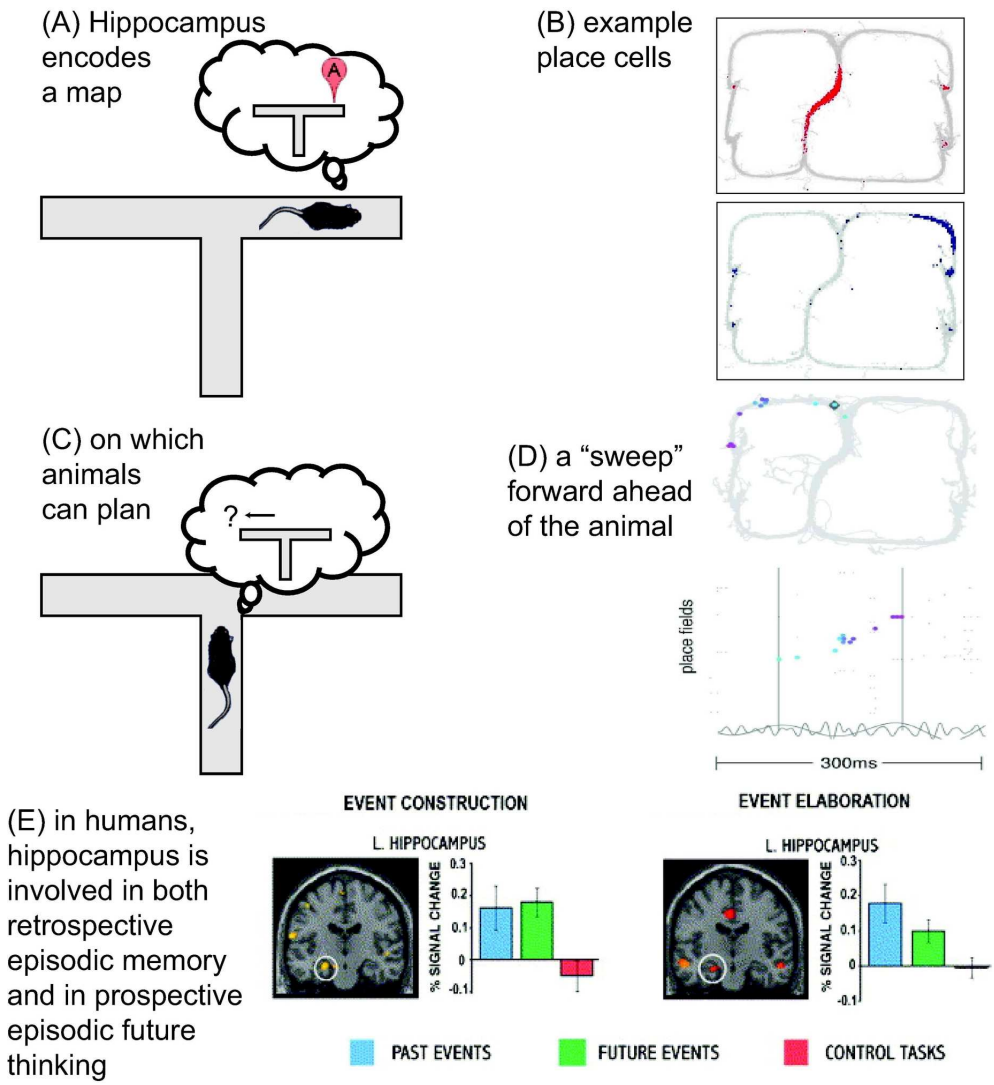
Pavlovian decision-making systems
169x184mm (300 x 300 DPI)



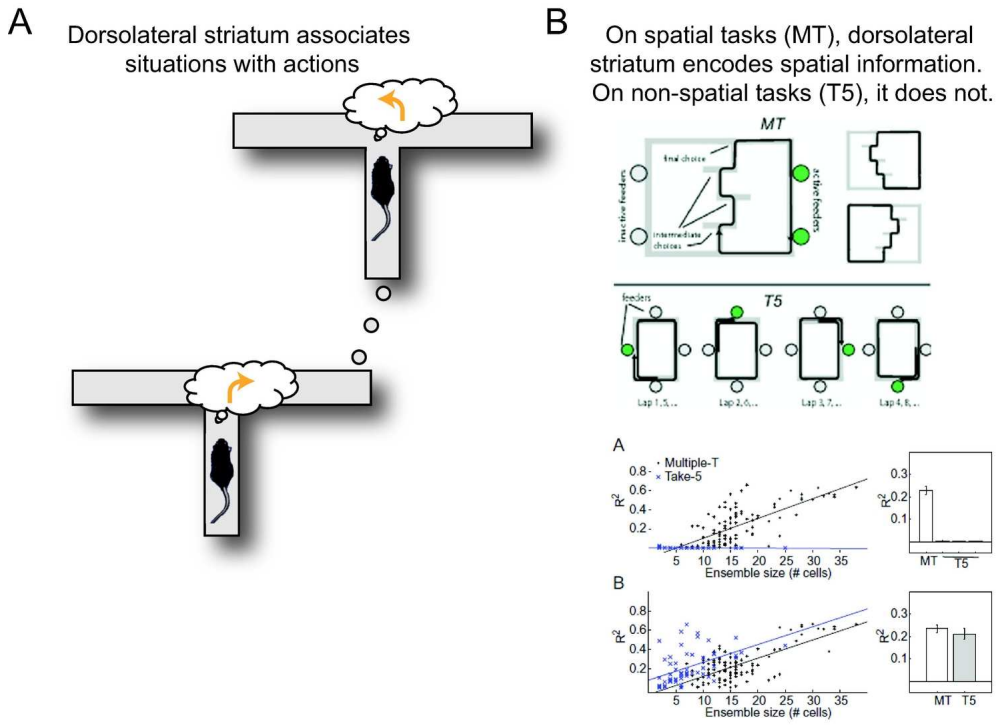
Habit-based decision-making
162x171mm (300 x 300 DPI)



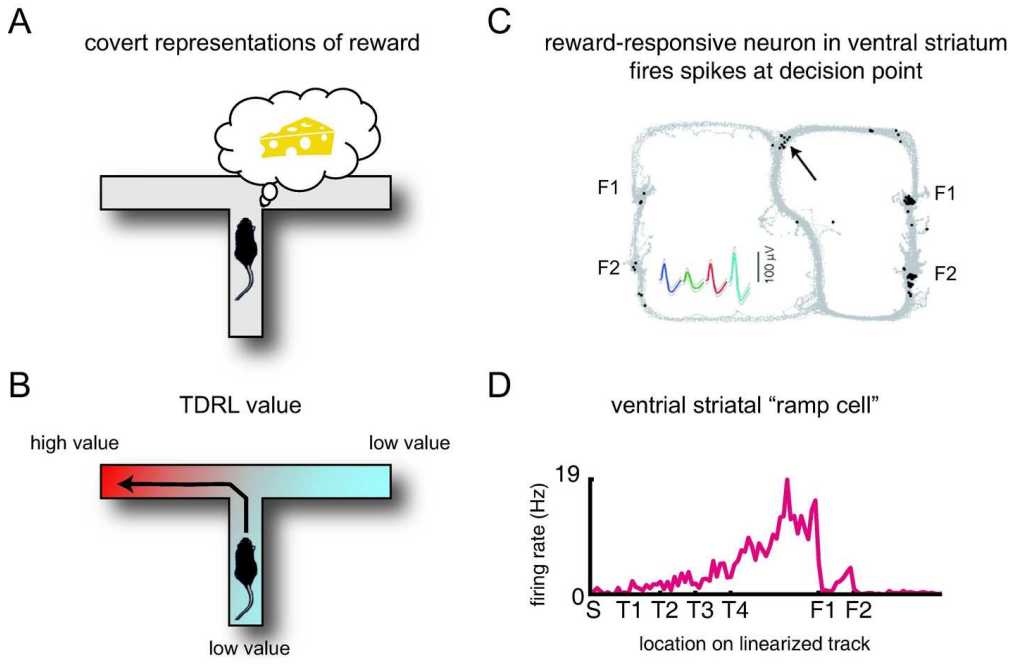
Deliberative decision-making
171x181mm (300 x 300 DPI)



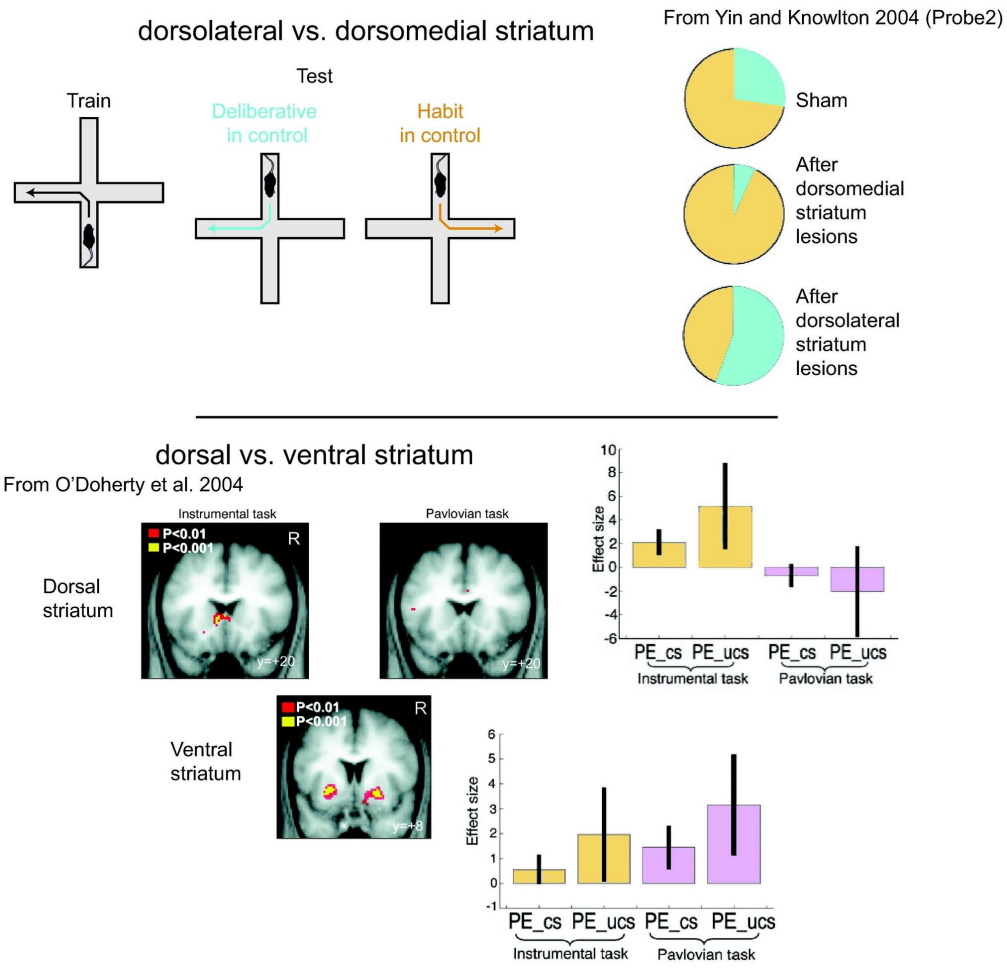
Hippocampal contributions to decision-making
158x172mm (300 x 300 DPI)



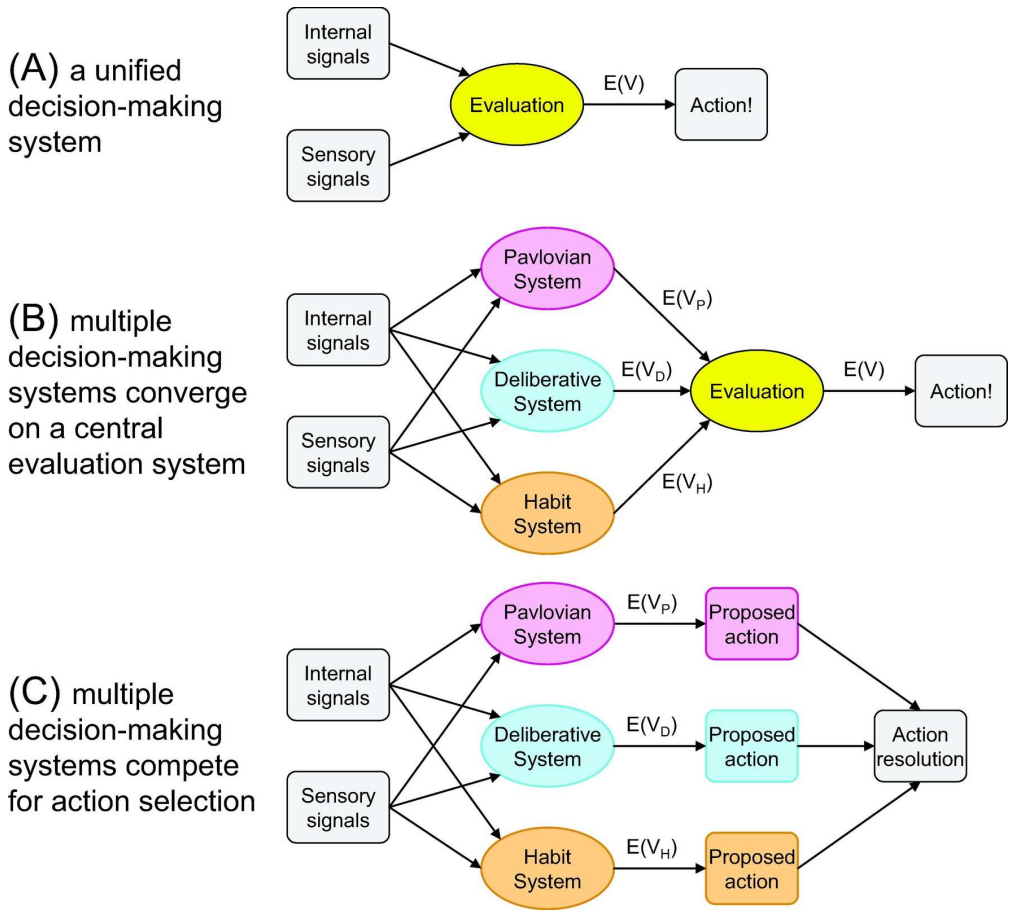
Dorsal Striatal contributions to decision making
126x91mm (300 x 300 DPI)



Ventral Striatal contributions to decision making
120x80mm (300 x 300 DPI)



Striatal components and decision making
170x163mm (300 x 300 DPI)



Reconciling multiple decision-making systems with neuroeconomic theories
149x135mm (300 x 300 DPI)