

**COMMENTARY**

# Learning, memory and consolidation mechanisms for behavioral control in hierarchically organized cortico-basal ganglia systems

Silviu I. Rusu<sup>1,2</sup> | Cyriel M. A. Pennartz<sup>1,2</sup> 

<sup>1</sup>Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup>Research Priority Program Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands

**Correspondence**

\*Cyriel M. A. Pennartz, Science Park 904, 1098 XH, Amsterdam, The Netherlands. Email: c.m.a.pennartz@uva.nl

**Funding information**

European Union Horizon 2020 Program and Netherlands Organization for Scientific Research (NWO), Grant/Award Numbers: 720270, 785907, 823.02.020

**Abstract**

This article aims to provide a synthesis on the question how brain structures cooperate to accomplish hierarchically organized behaviors, characterized by low-level, habitual routines nested in larger sequences of planned, goal-directed behavior. The functioning of a connected set of brain structures—prefrontal cortex, hippocampus, striatum, and dopaminergic mesencephalon—is reviewed in relation to two important distinctions: (a) goal-directed as opposed to habitual behavior and (b) model-based and model-free learning. Recent evidence indicates that the orbitomedial prefrontal cortices not only subserve goal-directed behavior and model-based learning, but also code the “landscape” (task space) of behaviorally relevant variables. While the hippocampus stands out for its role in coding and memorizing world state representations, it is argued to function in model-based learning but is not required for coding of action–outcome contingencies, illustrating that goal-directed behavior is not congruent with model-based learning. While the dorsolateral and dorsomedial striatum largely conform to the dichotomy between habitual versus goal-directed behavior, ventral striatal functions go beyond this distinction. Next, we contextualize findings on coding of reward-prediction errors by ventral tegmental dopamine neurons to suggest a broader role of mesencephalic dopamine cells, viz. in behavioral reactivity and signaling unexpected sensory changes. We hypothesize that goal-directed behavior is hierarchically organized in interconnected cortico-basal ganglia loops, where a limbic-affective prefrontal-ventral striatal loop controls action selection in a dorsomedial prefrontal–striatal loop, which in turn regulates activity in sensorimotor-dorsolateral striatal circuits. This structure for behavioral organization requires alignment with mechanisms for memory formation and consolidation. We propose that frontal corticothalamic circuits form a high-level loop for memory processing that initiates and temporally organizes nested activities in lower-level loops, including the hippocampus and the ripple-associated replay it generates. The evidence on hierarchically organized behavior converges with that on consolidation mechanisms in suggesting a frontal-to-caudal directionality in processing control.

**KEYWORDS**

hippocampus, model-based learning, nucleus accumbens, prefrontal cortex, striatum, ventral tegmental area

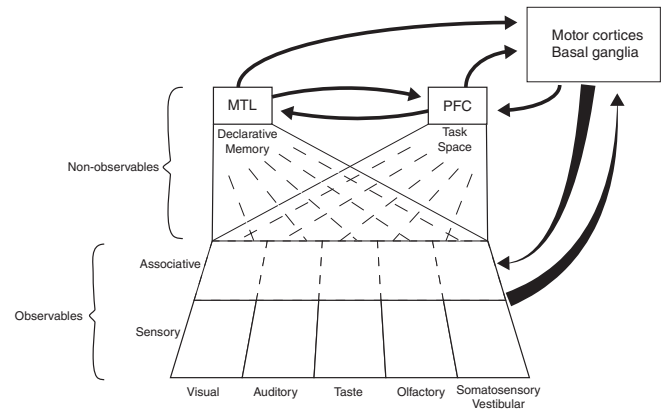
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Hippocampus* published by Wiley Periodicals, Inc.

## 1 | INTRODUCTION: WORLD MODELS OF OBSERVABLE AND NONOBSERVABLE VARIABLES

The idea that much of the neocortex is concerned with generating a world model, subserving decision-making and action, has gained much prominence and support in recent years. Here, the concept of “world model” is taken widely, including not only the representation of objects and their spatiotemporal context but also their statistical and causal relationships. The construction of a world model depends on the inference of the causes of sensory inputs the brain receives (Friston, 2010; Lee & Mumford, 2003; Pennartz, 2018), culminating in conscious perception set in different sensory modalities (Figure 1; Friston, 2010; Lee & Mumford, 2003; Pennartz, 2018). Because the total influx of sensory information is limited and partially incomplete, and may contain conflicting elements, this process amounts to making a “best guess” representation of the agent’s sensory world, which includes its own body (Friston, 2010; Gregory, 1980; Lee & Mumford, 2003; Marcel, 1983; Pennartz, 2015). Next to modeling the causes of environmental inputs, which reach the brain via sensory activation, a different component of world-modeling addresses latent or non-observable causes of events and situations. This type of cause cannot be directly verified by acute sensory input or motor behavior via which novel inferential information can be gained about sensory sources. In this context, “verification” reflects the process that a motor action, prompted by a particular sensory input (e.g., seeing a food item at a short distance, and reaching out to it), will result in sensory feedback through one or more other modalities (e.g., touching and tasting the object), which may turn out to be aligned with the initial visual estimate or not. Instead, nonobservable causes must be derived from long-term exploration of the environment, including manipulation of its many state variables, resulting in the discovery of causal relationships between relevant elements. This more long-lasting exploration is thought to result in a model laying out how specific hidden causes are related to each other as well as to observable effects. This more abstract model of nonobservable variables is as important for guiding behavior as perceptual, experiential representations are (Figure 1).

A key issue in building world models of hidden causes is that organisms must learn how their goals can be best achieved. The general goal of satisfying the organism’s homeostatic variables (in terms of survival and reproduction) can be translated into concrete situational needs such as food, water, sex, and avoidance of pain or stress. A classic paradigm for learning to optimize homeostatic conditions through behavior is reinforcement learning (RL), whereby the agent learns to couple stimuli to those actions resulting in maximal reward and minimal punishment (Barto, 1994; Pennartz, 1996; Sutton & Barto, 1998). During classic RL, stimuli and actions come to be associated with a cached value (i.e., a scalar value), accumulated over a long history of prior experiences, that predicts the value of the outcome of situations (as defined by, for example, reward magnitude). By consequence, learning processes become dependent not only on absolute



**FIGURE 1** Hierarchical organization of the mammalian brain defined by the progressive processing of observables and nonobservables and the interaction with brain structures for behavioral control and planning. Primarily unisensory areas are represented at the bottom row, feeding information to associative cortices for multisensory, higher-order representations of observables such as visible and audible objects. This information is propagated into medial temporal lobe (MTL) and prefrontal (PFC) structures. The memory system of the MTL is involved in computing nonobservables, such as the subject’s position in time and space, semantic meaning, name, and history of observed objects. The prefrontal cortex is proposed to encode a task space specifying relationships between task variables such as cues, actions, policies, outcomes, and motivational factors such as related to hunger or thirst. The interactions between PFC and the motor cortices and basal ganglia are expanded upon in Figure 2. In addition, the motor structures maintain bidirectional interactions with sensory and associative cortical areas. Note that not all known anatomical projections are included (e.g., from motor cortices to MTL). This scheme differs from previous proposals on hierarchical brain organization, for instance as proposed by Fuster (2001), who suggested two parallel hierarchies of sensory and motor processing streams converging upon the PFC as the highest center for integration. Note the following specific aspects of the current proposal: (a) the task space, encoding nonobservable relationships, has a rather abstract nature which is neither purely sensory nor motor; (b) The PFC and MTL, including the hippocampal formation, communicate intensively to plan and guide behavioral control; (c) the PFC and more caudally located motor-related areas form a system of hierarchically organized loops with the basal ganglia, which are also fed by information from MTL areas such as HPC and amygdala

reward, but also on differences between expected and actual reward (Rescorla & Wagner, 1972; Sutton & Barto, 1998).

More recently, behavioral control based on cached value has been contrasted to “model-based RL”, in which the agent builds an explicit, internal model of its state space, containing specific stimulus–outcome and action–outcome relationships. It is this latter type of sensory-specific learning that is referred to in the characterization of world models given above. Whereas classic RL guides actions based on a single scalar value reflecting reward history, model-based RL enables prospective cognition. It endows the agent with the capacity to predict specific future states, usually rendered as a decision tree with expanding ramifications (Daw & Dayan, 2014; Daw, Niv, &

Dayan, 2005). Because an agent can rely on its internal model when facing a choice situation, it is able to improvise or respond on the fly, using its general knowledge of relationships between specific stimuli, actions, and outcomes. It can flexibly respond to novel situations based on general knowledge laid down in tree structures and can make specific predictions about the immediate outcomes of each action alternative by associatively “chaining” short-term predictions. The internal model can be conceived as the core of a larger spatiotemporal model of causal structures of the world and the agent's position in it.

Model-based learning (MBL) is intimately related to, although arguably different from, the concept of goal-directed behavior (GDB). GDB does not refer just to any type of “goal-directed movement” (such as saccades), but to actions which are initiated based on representational content to control behavior (e.g., the belief that an action  $A$  is causal in obtaining outcome  $O$ ; Dickinson, 2012). Experimentally, GDB is assessed by testing whether a learned action is sensitive to outcome value, and whether animals learn that obtaining the outcome is contingent upon the action being performed (Balleine & Dickinson, 1998; Holland, 2004). In case neither of these criteria is met, actions are considered to result from stimulus–response learning followed by habit formation, which renders the execution of actions largely outcome-insensitive.

Because classic RL is devoid of a specific causal model, it has also been labeled “model-free learning” (MFL; Daw et al., 2005; Doya, 1999). In MBL, referral to a tree structure (or tree search) is motivated by the associative chains of short-term predictions being represented by an architecture that temporally unfolds as a branching set of possible outcome situations, in which an initial state  $S_0$  gives rise to multiple possible future states (e.g.,  $S_1$  and  $S_2$ ; Daw et al., 2005).

Several *operational indicators* for characterizing neural systems involved in MBL versus MFL can be delineated. An MBL system will be sensitive to *specific* properties of the outcome of an action in a specific context (Daw et al., 2005; Jones et al., 2012). In contrast, MFL depends on a slow accumulation of reward value over time and results in habitual behavior (HB) that is less flexible and less susceptible to readjustment. In MBL, the outcome is not only specified in terms of motivational value, but also by the specific sensory features defining its identity (e.g., apple flavor of a reward; Daw et al., 2005). Whereas for MFL only the value of an outcome matters, internal models have space for coding of sensorily distinct outcomes, even if they have the same value. An animal may be equally motivated to pursue a banana versus apple reward, but can nonetheless distinguish which of these equally valued outcomes will be obtained given an action  $X$  in situation  $Y$ .

Importantly, this does not imply that MFL systems do *not* distinguish between outcome-predicting actions or stimuli. Reward (or  $Q$ -) values will depend on the action chosen in a particular state  $S$  or context, even though only one scalar value is associated with a state-action pair (Daw et al., 2005). A second hallmark of MBL is its prospective and on-the-fly nature, going beyond the single value-prediction emitted in MFL (Daw et al., 2005; Doya, 1999). Neurophysiologically, this hallmark can be studied via the neural coding of

potential—but specific—choices or paths an animal may undertake or consider for future action, either during ongoing behavior, resting, or sleep.

Replay is the experience-dependent recurrence of temporally ordered sequences of neural activity characteristic of a preceding behavioral episode. In the hippocampus (HPC), this is expressed by firing patterns of neuronal ensembles coding sequences of sensory-specific states and is therefore richer, and more compatible with tree search, than would be expected from an integration across reward history such as in MFL. Moreover, replay may subserve both retrospective and prospective cognition, as illustrated in a spatial memory task (Jadhav, Kemere, German, & Frank, 2012). However, in applying these indicators for MBL versus MFL, it should be noted that they may not be uniquely characteristic for MBL but may also be compatible with other computational functions, and therefore do not offer proof for MBL per se.

Despite the close relationships between MBL and GDB, we argue that there are some conceptual and operational differences. First, whereas MBL originated from computational modeling, GDB was conceptualized to explain experimental observations on behavioral phenomena (outcome devaluation and contingency). Second, MBL and MFL are primarily concerned with particular structures and contents of what is learned and stored in memory. Although both types of learning subserve optimal action control, they do not specify how exactly actions are selected based on various neural systems. For instance, whether MBL and MFL systems cooperate and/or compete with each other remains a matter of debate (cf. Pezzulo, van der Meer, Lansink, & Pennartz, 2014). The MBL–MFL categorization does not specify computational mechanisms of directed action control or the precise neural substrates that play a role in them. Here, MBL and MFL differ conceptually from GDB versus HB, which are operationally defined via behavioral actions and outcomes, not as types of learning based on a particular structural model (e.g., a tree; Daw & Dayan, 2014). GDB hinges on the immediate importance of outcome value in guiding subsequent behavior and the causal importance of action–outcome relationships. While the concept of GDB does recognize the importance of representational content to control behavior (Dickinson, 2012), it does not specify how sensory-specific an outcome representation would have to be (regardless of value), whereas sensory specificity is a hallmark of MBL. Third, in contrast to MBL, GDB is not defined by capacities for retrospective or prospective cognition or on-the-fly improvisation, although it is certainly compatible with such processes.

In addition to learned behaviors, reflexes and innate behaviors need to be considered in the optimization of homeostatic variables. However, due to space constraints, these largely fall outside the scope of the current review. The same is the case for learning paradigms that do not inherently contain a definition of the outcome (such as associative object-place learning or socially transmitted food preference) and hence are compatible with either MBL or MFL. The current focus on the MBL–MFL and GDB–HB distinctions does not preempt other, often more detailed proposals for decision-making; our focus is rather global and involves multiple connected brain systems. As such,

it is broadly compatible with more detailed schemes that often zoom in on particular brain structures or systems (e.g., basal ganglia, Pennartz, Groenewegen, & Lopes da Silva, 1994; Redgrave, Prescott, & Gurney, 1999a; Wei & Wang, 2016; lateral intraparietal area [LIP], Shushruth, Mazurek, & Shadlen, 2018; orbitofrontal cortex [OFC], Padoa-Schioppa & Conen, 2017).

While much remains unknown about the neural substrates of MBL versus MFL, there has been a tendency in the literature to associate MFL with the dorsal and ventral basal ganglia (Houk, Adams, & Barto, 1995; Schultz, 1998) and MBL with prefrontal cortex (PFC; Daw & Dayan, 2014; Daw et al., 2005; Jones et al., 2012). Within neuroscience, temporal difference reinforcement learning (TDRL; Sutton & Barto, 1998) has gained much prominence as an effective and plausible type of MFL. In mapping an actor-critic architecture that implements TDRL onto the basal ganglia, the dorsal striatum would function as "actor," mediating actions that impinge on the subject's environment to evoke sensory and reinforcing feedback. In contrast, the ventral striatum (VS) would be the "Critic" that generates outcome-predicting values based on the history of reinforcement obtained from these actions (Houk et al., 1995; Schultz, 1998). The ventral tegmental area (VTA) would compute errors between predicted and actual reward (reward prediction errors), and transmit these to the Actor module to guide learning of stimulus and action values, resulting in the selection of the most valued action.

Finally, the concept of hierarchical control needs some further introduction in the context of GDB and HB. GDB can be usually decomposed into a sequence of subroutines or skills carried out to reach an end means, such as an animal locomoting to a lever, pressing the lever, and moving over to a magazine site to ingest and swallow a food pellet. Classically, the analysis of sequential, goal-directed behavior has supported the tenet that it cannot be understood as a simple chain of stimulus-response associations, but is characterized by a hierarchical organization in which low-level subroutines are organized to subserve attainment of the end goal (Lashley, 1951; Miller, Galanter, & Pribram, 1960). This analysis was later supported by the notion that actions are not only controlled by cues acutely available to subjects (e.g., a light prompting a rat to press a lever for reward) but also by the wider spatiotemporal context in which actions are performed (e.g., the reward can only be obtained in a particular environment or after a chain of events). Thus, a higher-order system is needed to consider options for actions in the light of the supraordinate context (Badre & Nee, 2018; Desrochers, Chatham, & Badre, 2015). A further, computational argument for hierarchical control, which arose from studies on RL, is that RL models become less effective when coping with many possible actions and world states. This "curse of dimensionality" can be mitigated by grouping small-scale actions together into more abstract, temporally extended actions (e.g., "perform the lever press sequence" in the example above). This computational problem has been a driving force for developing hierarchical RL models, conforming to the idea that complex behavior comprises both elementary actions and overarching action patterns (Barto & Mahadevan, 2003; Botvinick, 2008; Chiang & Wallis, 2018; O'Reilly & Frank, 2006). Thus, behavioral and computational

arguments support the notion of hierarchical control over behavior, which however does not imply that the underlying neural substrates necessarily have an explicitly hierarchical structure (Cleeremans, Destrebecqz, & Boyer, 1998; Elman, 1990).

It is in this theoretical and empirical context that we will review the evidence for the variously proposed roles of the PFC, HPC, and connected basal ganglia structures in GDB versus HB and MBL versus MFL, and in hierarchically organized behavior. Besides lesion and other interventional studies, we will emphasize electrophysiological studies on neural coding in these structures in the rodent brain. The review is structured as follows. First, we discuss evidence for roles of particular brain areas (and their main subdivisions) in MBL versus MFL and GDB versus HB, in relation to other functions associated with the same areas. Next, we consider how these structures are jointly organized in interacting cortico-basal ganglia loops to provide a plausible neural substrate for a hierarchical organization of behavior, in which systems for GDB and MBL assume a higher position than systems for HB and MFL. Finally, we review communication mechanisms in cortico-basal ganglia systems and ask whether "offline" processing in corticothalamic and hippocampal circuits aligns with the proposed frontal-to-caudal direction in behavioral hierarchical control.

## 2 | PREFRONTAL CORTEX: MODEL-BASED LEARNING, GOAL-DIRECTED BEHAVIOR, AND THE CODING OF TASK SPACE

### 2.1 | Introductory remarks

The question of whether the PFC is involved in GDB, MBL, or their counterparts is positioned in a rich literature having raised evidence for its role in a gamut of cognitive functions, which includes working and long-term declarative memory, categorization, decision-making, cognitive flexibility, attentional shifting, outcome valuation, control of emotions, and self-initiation of behavior (Bari & Robbins, 2013; Eichenbaum, 2017; Fuster, 2001; Goldman-Rakic, 1996; Miller & Cohen, 2001; Pennartz, van Wingerden, & Vinck, 2011). This diversity of functions may to some extent be attributable to distinct PFC subregions playing different roles. For instance, the medial PFC (mPFC) has been implicated in flexible rule learning and working memory (Euston, Gruber, & McNaughton, 2012; Mulder, Nordquist, Orgut, & Pennartz, 2003; Rich & Shapiro, 2009; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011; Wallis, Anderson, & Miller, 2001), the anterior cingulate cortex in detecting response conflict and decision-making based on effort and temporal cost-benefit constraints (Cowen, Davis, & Nitz, 2012; Haddon & Killcross, 2006; Hosokawa, Kennerley, Sloan, & Wallis, 2013; Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006, but see Walton, Croxson, Behrens, Kennerley, & Rushworth, 2007), and the OFC in flexibly learning stimulus-outcome value (Jones et al., 2012; Ostlund & Balleine, 2007; Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009; van Duuren, Lankelma, & Pennartz, 2008; van Duuren et al., 2009; van Wingerden et al., 2012; but see Stalnaker, Cooch, & Schoenbaum, 2015). In this sense, one applies a

restricted lens on PFC when considering it in the light of GDB and MBL. Below we will treat the roles of OFC and mPFC separately.

## 2.2 | Orbitofrontal cortex: model-based versus model-free learning and goal-directed versus habitual behavior

While regional differences may account for part of the observed diversity in PFC functions, they also raise the question whether all or most of these functions can be subsumed under a common functional denominator, such as *MBL as opposed to MFL*. It could be argued that many results obtained from lesion and physiological studies, pertaining to decision-making, cognitive flexibility, and outcome valuation may all be captured under the common framework of MFL. Confirming to this, the OFC has been proposed to fulfill the role of “critic” based on results from lesion and electrophysiological studies showing that this structure plays a major role in coding and applying stimulus-value associations, in conjunction with the amygdala (Ostlund & Balleine, 2007; Pennartz, Ito, Verschure, Battaglia, & Robbins, 2011; Schoenbaum, Setlow, Saddoris, & Gallagher, 2003). Single units and ensembles in OFC display learning-dependent, predecisional responses to odors or other outcome-predicting stimuli (Schoenbaum et al., 2003; van Wingerden et al., 2012), correlating with the magnitude or probability of upcoming reward (van Duuren et al., 2008; van Duuren et al., 2009). Reward value is not coded by OFC neurons generally, but is often coupled to a specific sensory stimulus, such as one of several odors that predict reward (Ramus & Eichenbaum, 2000; van Duuren et al., 2008). By itself, however, such sensory specificity is compatible with both MBL and MFL.

However, there is considerable evidence that an MFL-based scheme does not capture the full complexity of OFC coding. In a sensory preconditioning paradigm using muscimol and baclofen to inactivate the OFC, Jones et al. (2012) found that the OFC is needed for both MBL and MFL, but is not critically involved when cached value is sufficient for decision-making. This conclusion could be drawn because, in this task, the value of a secondary cue, not paired with reward, had to be inferred from its pairing to a primary, reward-paired cue (McDannald et al., 2012; Stalnaker et al., 2015). This result does not imply that the OFC is not involved in MFL but does allow us to conclude that its role goes further, using inferential representations of environmental structure. This proposal gains further support from the relatively fast, plastic changes in OFC firing responses to olfactory stimuli during reversal learning (van Wingerden, Vinck, Lankelma, & Pennartz, 2010a; cf. Burke, Takahashi, Correll, Brown, & Schoenbaum, 2009). Such fast changes would not be expected if the OFC would only support MFL (cf. Daw et al., 2005). In addition, lesion studies have implied the OFC in reversal learning (e.g., Rolls, 2000; but see Stalnaker et al., 2015).

Moreover, recent neurophysiological evidence supports a function of the OFC in MBL. When an animal has been trained to run a maze and make wait or skip choices for delayed delivery of differently flavored food pellets at distinct maze sites, a regret condition occurs when the animal commits to a high-cost choice after having skipped a

low-cost option. In this situation, orbitofrontal (and ventral striatal) ensembles strongly represented the previous low-cost option after the rat had entered the current, high-cost zone (Steiner & Redish, 2014). Because this coding pertains to a previously encountered choice in the decision tree and indicates a form of retrospective cognition, this result is in line with the proposed role for OFC in MBL, while not refuting an additional role in MFL.

As concerns the role of OFC in *GDB* versus *HB*, evidence for its causal involvement in GDB has been mounting in recent years. In marmoset monkeys, Jackson, Horst, Pears, Robbins, and Roberts (2016) found that OFC lesions caused an insensitivity to degradation of action–outcome contingency, underpinning an important component of GDB (see Valentin, Dickinson, & O’Doherty, 2007, for fMRI results in humans). In mice subjected to a within-subject lever-pressing task using reinforcer devaluation, chemogenetic inhibition of OFC disrupted GDB, while optogenetic activation increased this behavior (Gremel & Costa, 2013). In rats, learning instrumental actions in the setting of an outcome devaluation paradigm, chemogenetic inhibition of the insular cortex inhibited goal-directed control, while inhibition of the ventrolateral OFC also impaired GDB, albeit only after a switch in instrumental contingencies (reversal training; Parkes et al., 2018).

## 2.3 | Medial prefrontal cortex: model-based versus model-free learning and goal-directed versus habitual behavior

Although some studies refer to the mPFC as a neural substrate for MBL based on arguments from action contingency or devaluation, it is important to determine whether the experimental evidence supports a role specifically in MBL or GDB. While the evidence for GDB is significant (see below), there are some studies pointing to a role in *MBL*, as opposed to *MFL*, as well. In a value-based decision task in which forward planning was contrasted with cache-based choices acquired through extensive training, Wunderlich, Dayan, and Dolan (2012) found that BOLD activity in the human dorsomedial PFC, along with other structures such as the medial frontal gyrus and precuneus, was enhanced during planning relative to cache-based trials. Similar studies contrasting MBL versus MFL found BOLD correlates in additional brain structures well connected to mPFC, such as the intraparietal sulcus and lateral PFC (Gläscher, Daw, Dayan, & O’Doherty, 2010) or medial temporal lobe and striatum (Simon & Daw, 2011). These findings are generally in line with models implying the mPFC in counterfactual reasoning and future planning (Barbey, Krueger, & Grafman, 2009).

Electrophysiological studies highlight that mPFC firing patterns are often highly specific for particular action sequences, run paths, goal sites, and memory strategies (Euston et al., 2012; Hok, Save, Lenck-Santini, & Poucet, 2005; Ito, Zhang, Witter, Moser, & Moser, 2015; Kargo, Szatmary, & Nitz, 2007; Mulder et al., 2003; Rich & Shapiro, 2009). This sequence specificity is also shown in replay generated by mPFC assemblies during sleep or behavioral pausing (Euston, Tatsuno, & McNaughton, 2007). Evidence for a causal involvement of mPFC in prospective, MBL-based cognition relates to its influence on

vicarious trial and error (VTE) behavior and the associated process of “forward sweeps” (theta sequences) in hippocampal representations generated when animals are at a choice point in a maze. Schmidt, Duin, and Redish (2019) found that disruption of mPFC activity using DREADD manipulation diminished VTE behavior in a spatial foraging task and impaired theta-sequence generation in area CA1. Altogether, the evidence supports a causal role of mPFC in MBL, although sensory-specific aspects of outcome coding remain to be tested by single-unit recordings.

A causal role of mPFC in GDB, as opposed to HB, has been suggested by Balleine and Dickinson (1998). Their paradigm tested for specific actions (lever pressing vs. chain pulling) being coupled to specific outcomes (food pellets vs. starch solution) and showed that the outcome sensitivity of action learning was dependent on mPFC integrity (Balleine & O’Doherty, 2010; Corbit & Balleine, 2003). Lesions of the mediodorsal thalamic nucleus, which is bidirectionally connected with the PFC, also degraded GDB as measured by action–outcome devaluation (Corbit, Muir, & Balleine, 2003). Coutureau and Killcross (2003) suggested that the prelimbic and infralimbic PFC subregions play differential roles in GDB versus HB, as muscimol inactivation of infralimbic cortex allowed animals to reinstate goal-directed responding following extended training, whereas control-infused animals continued to show HB (see, however, Shipman, Trask, Bouton, & Green, 2018).

## 2.4 | Further analysis: Prefrontal coding of task space

Apart from the MBL–MFL distinction, additional observations suggest that mPFC function is more comprehensive than simply coding specific action–outcome relationships. Earlier studies noted that all elements relevant to task performance are represented in firing correlates of mPFC neurons, including trial-initiation stimuli and stimuli coupled to contingent action and reward (e.g., Baeg et al., 2003; Kargo et al., 2007; Mulder et al., 2003). Because this sequential coding of task elements (“tessellation”; Pennartz, van Wingerden, & Vinck, 2011) includes a substantial component of stimulus and context information, the mPFC likely encodes task stages preceding the action–outcome phase. This tessellation has also been found in other cortical and subcortical areas (Allen et al., 2017; Bos et al., 2017; Harvey, Coen, & Tank, 2012; Lansink, Goltstein, Lankelma, & Pennartz, 2010). As regards mPFC, however, lesion and electrophysiological studies are consistent with a causally relevant role in coding behaviorally relevant stimuli and contexts (Birrell & Brown, 2000; Euston et al., 2012; Mulder et al., 2003; Takehara-Nishiuchi & McNaughton, 2008).

Another indication for a broader repertoire of PFC functions came from studies showing coding of task rules and goals in this structure in macaques and rodents (notably, this coding is found in medial, but also lateral and ventral parts of PFC; Durstewitz, Vittoz, Floresco, & Seamans, 2010; Genovesio, Tsujimoto, & Wise, 2012; Wallis et al., 2001). These rules amount to if–then relationships applicable in a specific condition (e.g., “if stimulus X appears in situation Y, then perform action A; if X appears in situation Z, perform action B”). In a learning

paradigm including multiple, serial reversals, De Bruin et al. (2000) found that lidocaine infusions in mPFC transiently impaired the first instance of reversal learning, suggesting that the mPFC is required for fast instatement of new task rules when expected outcomes are no longer obtained. These and other findings gave rise to the concept of PFC as coding a “task space” (Verschure, Pennartz, & Pezzulo, 2014). In contrast to geometric spaces, this concept holds that the PFC codes an abstract map of the nonobservable, causal relationships between task elements, as far as relevant for achieving end goals. The coding of task space may result from MBL and acts as an informational reservoir to drive GDB. A major outstanding question is whether the PFC stores task-space information in its synaptic matrices, or mainly imports the information from other areas and utilizes it to compute its outputs on-line, thus steering behavior (cf. Jones et al., 2012).

In summary, current evidence indicates that coding in the orbitofrontal and medial prefrontal cortices goes well beyond MFL and may be more adequately captured by MBL, while not excluding value coding according to cache-based schemes. Similarly, experimental evidence supports a role for mPFC, OFC, and adjacent areas (e.g., insular cortex) in GDB. In addition, the functions of these cortices are not restricted to action–outcome relationships, but rather comprise the full “landscape” (task space) of stimuli, contexts, rules, actions, and outcomes as far as relevant for reaching goals. Within this task space, additional functions going beyond the MBL–MFL distinction are expressed by PFC ensembles, such as working memory and attentional set-shifting.

## 3 | HIPPOCAMPUS: MODEL-BASED LEARNING, GOAL-DIRECTED BEHAVIOR, AND THE MAPPING OF WORLD-STATE VARIABLES

### 3.1 | Introductory remarks

Hippocampal function has been the subject of excellent recent reviews (e.g., Buzsáki & Moser, 2013; Eichenbaum, Sauvage, Fortin, Komorowski, & Lipton, 2012), therefore we will restrict this section mainly to aspects relevant for the MBL–MFL and GDB–HB distinctions. First, we recall that hippocampal lesions primarily cause deficits in types of cognition not specifically associated with RL, viz. in spatial memory (Handelmann & Olton, 1981; Morris, Anderson, Lynch, & Baudry, 1986) and other aspects of episodic memory (Corkin, 2002; Milner, Squire, & Kandel, 1998; Tulving, 1983). In rodents, hippocampal lesions also affect working memory and time-sensitive forms of conditioning (e.g., trace conditioning; Meck, Church, & Olton, 2013; Weiss & Disterhoft, 2015); working memory is defined here as a short-term form of memory holding information and allowing its manipulation for subsequent behavioral decisions. There is little evidence to implicate the HPC in processes lying at the core of RL, such as Pavlovian conditioning (without a delay component) and learning of singular stimulus–response associations (S–R learning), consistent with the notion that, in general, procedural memory and classical conditioning are not affected by hippocampal lesions.

Before reviewing hippocampal functions in MBL, GDB, and their counterparts, we will first examine the range of parameters that the HPC codes, such as a subject's spatial location in its environment (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978; Wilson & McNaughton, 1993). This range has been recently broadened to include the coding of time (Eichenbaum, 2014; Kraus, Robinson 2nd, White, Eichenbaum, & Hasselmo, 2013) and other behaviorally relevant, parametric variables. Depending on task design, hippocampal cells code the time elapsed between task-relevant events such as odor and object presentations, or the time spent during treadmill running (MacDonald, Carrow, Place, & Eichenbaum, 2013; MacDonald, Lepage, Eden, & Eichenbaum, 2011; cf. Pastalkova, Itskov, Amarasingham, & Buzsáki, 2008). Although the HPC was already proposed to represent nonspatial parameters in the 1980s (e.g., Meck, 1988; Weiss & Disterhoft, 2015), this concept recently gained strength by studies reporting hippocampal coding of temporal sequences during task execution. A causal role of the HPC in representing sequences of nonspatial events (i.e., odor choices) was indicated by a lesion study (Agster, Fortin, & Eichenbaum, 2002). Allen, Salz, McKenzie, and Fortin (2016) used an olfactory task in which rats were required to sample a sequence of odors and to identify each odor being presented as part of a sequence or "out of sequence." This study showed that dorsal CA1 cells code for multiple stimulus parameters, including stimulus identity, stimulus rank, or rank identity associations.

A recent study by Aronov, Nevers, and Tank (2017) showed that hippocampal coding is not limited to unimodal representation of one parametric variable such as spatial position. When rats were trained to manipulate a joystick, which served to modulate tone frequency, and reach into a target zone of frequencies to obtain reward, they demonstrated the emergence of hippocampal "auditory maps." In this task, the firing of CA1 neurons mapped onto the entire range of presented sound frequencies, similarly to place cells during spatial tasks, with subsets of cells showing preferred sound frequencies. In addition, a recent study showed that also sequences of egocentric body movements can be coded by HPC. In a star-shaped maze, where spatial navigation was based either on external landmark configurations (place memory) or on memorized sequences of body turns, mouse hippocampal CA1 ensembles showed correlates not only of place, but also of specific components in the sequence of body movements (Cabral et al., 2014). In mice where NMDA receptors on CA1 pyramidal cells were genetically deleted, this body-sequence based mapping was selectively degraded, paralleled by a behavioral deficit in memory for longer motor sequences. Thus, depending on task requirements, the HPC can code a broad range of parameters, which however does not preclude a role in MFL.

### 3.2 | Role in model-based versus model-free learning and goal-directed versus habitual behavior

In general, the rich episodic nature of hippocampal coding, as demonstrated by lesion studies and electrophysiological recordings, supports a role in MBL, while not excluding an additional function in MFL. More

specific evidence for a hippocampal role in MBL has been raised by ensemble recording studies revealing a relationship between "forward sweeps" representing potential future trajectories and VTE behavior displayed by rats at decision points in a maze (see above; Johnson & Redish, 2007; Wikenheiser & Redish, 2015). Excitotoxic HPC lesions were shown to affect VTE behavior in a spatial memory task, whereas VTE behavior during a visual discrimination task was not deteriorated. In the spatial memory task, sham-lesioned animals showed more VTE behavior as long as the reward location had not been identified, in comparison to trials performed after it had been located, whereas this difference was not found in lesioned animals (Bett et al., 2012). Neural substrates of these and other forms of prospective sequences (Pfeiffer & Foster, 2013) have been hypothesized to serve as an internal, sampling-based mechanism for computing and evaluating potential paths toward goal sites, optimizing decision-making (Penny, Zeidman, & Burgess, 2013; Pezzulo et al., 2014; Redish, 2016; Stoianov, Pennartz, Lansink, & Pezzulo, 2018).

As concerns hippocampal functions in GDB versus HB, Corbit and Balleine (2000) showed that electrolytic hippocampal lesions had no effect on the sensitivity of rat instrumental performance to outcome devaluation, but did have a deteriorating effect on sensitivity to degradation of action–outcome contingencies. At first, this suggested that the HPC is important for representing action–outcome causality, but in a follow-up study, using excitotoxic lesions, this effect was attributed to the entorhinal cortex and its efferents to the retrohippocampal area (Corbit, Ostlund, & Balleine, 2002).

Despite this lack of conclusive evidence, there are physiological indications that hippocampal processing is compatible with a role in GDB. In rats navigating a Y-maze, guided by nine spatially distributed cue lights predicting reward, Lansink et al. (2016) investigated theta and beta (15–20 Hz) rhythmicity in firing patterns and local field potentials (LFPs) in area CA1. Importantly, rats were not only prompted by a light cue to approach these goal sites, but also visited them without cue, by way of habitual "checking." Theta and beta-band (15–20 Hz) rhythmicities were augmented during goal approach guided by cues, relative to habitual approaches.

### 3.3 | Further analysis: Hippocampal representation of world states

Given the evidence for a richer role of HPC than is captured by spatial coding, the question arises how this structure is distinct from other areas coding a similarly varied repertoire of behaviorally relevant variables, such as PFC. While HPC lesions do not disrupt behavior in general, they specifically impair those behaviors depending on episodic memory, that is, memory for objects and events, their spatial locations, temporal order, and times of occurrence (Eichenbaum et al., 2012; Eichenbaum, Dudchenko, Wood, Shapiro, & Tanila, 1999; Fortin, Agster, & Eichenbaum, 2002; Kesner, Hunsaker, & Warthen, 2008; Meck et al., 2013; Ranganath, 2010; Scoville & Milner, 1957). A key proposition holds that the HPC does not encode a "task space" as proposed for PFC, but a representation of world states (including the subject's body state; cf. Verschure et al., 2014). At first glance, the

HPC and PFC seem to code many of the same variables, but we argue that this similarity is superficial. For instance, a collection of hippocampal place cells codes for every location occupied by an animal in space, including relevant as well as irrelevant sites, whereas prefrontal ensembles more prominently code task-relevant elements, such as goals and larger spatial or temporal task segments leading up to goals (Genovesio et al., 2012; Hok et al., 2005; Mulder et al., 2003; Rich & Shapiro, 2009). This evidence is corroborated by lesion and pharmacological studies more generally implying PFC in executive functions (Dalley, Cardinal, & Robbins, 2004; Gläscher et al., 2012; Miller & Cohen, 2001).

This concept anchors the HPC more firmly to the memory of sensory, motor, and spatiotemporal variables and derived nonobservable variables such as allocentric position in space, and anchors the PFC to mapping the latent structure of relationships between task elements (in particular their causal relationships such as their instrumental role in achieving goals). Support for these differential roles comes from studies using mazes or other spatial environments, where dorsal CA1 neurons show small place fields scattered across the environment (Bos et al., 2017; Davidson, Kloosterman, & Wilson, 2009), whereas OFC, mPFC (and the interconnected perirhinal cortex, Bos et al., 2017) predominantly code large, task-bound stretches of space, demarcated by preceding and consecutive task stages requiring switches in behavior (Rich & Shapiro, 2009).

To conclude, hippocampal function is not only characterized by coding state variables, but also includes its operation as a sequence generator and repository for pointers to store and retrieve elements of episodic memory (cf. Teyler & DiScenna, 1986). By this retrieval, relational information can be quickly utilized to plan and organize GDB. The apparent contrast between hippocampal lesion studies—indicating no causal role in GDB—and neurophysiological studies may be explained by assuming that the HPC strongly supports GDB via its episodic memory capacities and internally generating prospective sequences, in particular by providing contextualized sensory and motor information to GDB systems. However, the HPC is not causally required to code action–outcome relationships themselves, which may be uniquely or redundantly coded by other brain areas such as mPFC. In this sense, a lack of evidence for a causal role in GDB does not imply a lack of involvement in MBL. This conclusion is in line with the rationale to maintain a conceptual distinction between GDB and MBL (see section 1).

## 4 | STRIATUM: MODEL-BASED LEARNING, GOAL-DIRECTED BEHAVIOR, AND THE CONVERSION FROM STATE TO ACTION INFORMATION

### 4.1 | Introductory remarks

There are several, mutually consistent grounds to argue that the notion of the dorsal and ventral striatum recalled above—as “Actor” and “Critic” in MFL—is not sufficient to capture the diversity and complexity of processing within this structure, if the concept is valid at all (Pennartz, Ito, et al., 2011; van der Meer & Redish, 2011b). Consistent

with the anatomically similar structure of cortico-basal ganglia–thalamic loops involving dorsoventral but also mediolateral gradients in the striatum (Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2004), lesion studies suggested that striatal sectors differ by the domain of information processing and learning, rather than by an actor–critic type of division (Pennartz, Ito, et al., 2011). The ventromedial striatum (nucleus accumbens shell, receiving substantial ventral hippocampal CA1–subicular inputs; Groenewegen, Vermeulen-Van der Zee, te Kortschot, & Witter, 1987) has been implicated in space–outcome and context–outcome associations (Ito, Robbins, Pennartz, & Everitt, 2008), whereas the ventrolateral sector (nucleus accumbens core, receiving strong amygdala inputs) is more clearly involved in cue–outcome learning and Pavlovian-to-instrumental transfer (PIT; Cardinal, Parkinson, Hall, & Everitt, 2002; Hall, Parkinson, Connor, Dickinson, & Everitt, 2001). Both shell and core receive converging afferents from mPFC and OFC, albeit in a subregion-specific manner (Pennartz et al., 1994; Pennartz, Ito, et al., 2011; Voorn et al., 2004). Similarly, the dorsal striatum has been subdivided into a dorsomedial and dorsolateral region (DMS and DLS) based on different anatomic input–output relationships and distinct functionalities. A comparison of electrophysiological studies studying neural coding in different striatal regions suggests prominent representation of cue value in rat VS and primate caudate (e.g., Kim & Hikosaka, 2013; Lansink et al., 2012; Roitman, Wheeler, & Carelli, 2005). Furthermore, information on task conditions (e.g., fixed versus free-choice trials) and state value is strongly represented in rat VS (Ito & Doya, 2015; Lansink et al., 2012; Lansink et al., 2016; Schultz, Apicella, Scarnati, & Ljungberg, 1992), whereas action value is more prominently coded in dorsal striatum (Ito & Doya, 2015; Samejima, Ueda, Doya, & Kimura, 2005). Within dorsal striatum, the lateral and medial sectors display different firing-rate dynamics and task correlates during learning (Thorn, Atallah, Howe, & Graybiel, 2010).

### 4.2 | Dorsal striatum: model-based versus model-free learning and goal-directed versus habitual behavior

The involvement of DMS and DLS in MBL versus MFL has been addressed by various methods. In a human fMRI study, Wunderlich et al. (2012) found evidence for MBL-related activity in the anterior caudate in the value-based forward planning task referred to above, whereas the putamen was implied in value representation acquired through extensive, MFL-related training. Using a spatial planning task, Simon and Daw (2011) found BOLD correlates of plan-based predicted values in VS but also putamen, arguing for a more widespread role of MBL in the striatum than previously thought. In ensemble recording studies, prospective coding was tested in a multiple T-maze task where rats navigated for reward, while recordings were made from HPC, dorsal, and ventral striatum. In contrast to the HPC and VS, the dorsal striatum did not show prospective representations of path options and reward, but displayed more pronounced coding of task-related actions, consistent with a stronger role of dorsal striatum in MFL (van der Meer, Johnson, Schmitzer-Torbert, & Redish, 2010).



In this study, DMS and DLS were not separately assessed. In an odor discrimination task for rats, neural coding of outcome specificity (identity), as dissociated from generic value, was surprisingly found in both DMS and DLS (Stalnaker, Calhoun, Ogawa, Roesch, & Schoenbaum, 2010).

The causal involvement of the dorsal striatum in *GDB* versus *HB* has been addressed in lesioning studies, implying the DMS in action–outcome learning (Hart, Bradfield, & Balleine, 2018; Hart, Leung, & Balleine, 2014; Yin, Ostlund, Knowlton, & Balleine, 2005), contrasting to the DLS, which has been implied in stimulus–response coupling and habit formation (Gremel & Costa, 2013; Pennartz et al., 2009; Smith & Graybiel, 2013; Yin, Knowlton, & Balleine, 2004). This dichotomy is supported by electrophysiological recordings reporting greater DMS engagement during goal-directed actions, as opposed to less engagement of the DLS (Gremel & Costa, 2013).

### 4.3 | Ventral striatum: model-based versus model-free learning and goal-directed versus habitual behavior

Evidence for a role of the VS in *MBL*—in addition to *MFL*—has been raised in several lesioning and electrophysiological studies. Excitotoxic VS lesions impaired both *MBL* and *MFL* in a Pavlovian blocking and unblocking paradigm, indicating the causal importance of VS in coding specific features of expected outcomes (McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011). Moreover, VS neurons respond differentially to stimuli that predict sensorily distinct, but equally or similarly valued rewards (Cooch et al., 2015; Gmaz, Carmichael, & van der Meer, 2018). In a task where rats ran a triangular track with distinct outcomes at spatially separated reward sites, Lansink et al. (2008) found cells that were either generally responsive to every site and every type of reward, or responded to reward at only one specific site. Following the hypothesis of distributed ensembles within the VS exerting different functions (Pennartz et al., 1994), these results indicate the presence of both generally and item-specific reward-predicting cell populations in VS. Further evidence for a VS function in *MBL* comes from the finding that VS cells emit spikes coding for expected reward along with look-ahead sequences in HPC (Pezzulo et al., 2014; van der Meer & Redish, 2009).

As regards *GDB* versus *HB*, the VS appears to be more involved with the motivational control of behavioral performance rather than with selecting goal-directed or habitual actions themselves (Hart et al., 2014). This is supported by excitotoxic lesion studies using PIT paradigms, showing that VS lesions disrupted gain modulation of instrumental action by Pavlovian stimuli, whereas the sensitivity to outcome devaluation or degradation of instrumental contingency was not affected, in contrast to DMS lesions (de Borchgrave, Rawlins, Dickinson, & Balleine, 2002; Hart et al., 2014).

### 4.4 | Further analysis: Ventral striatal transformation of spatial and cue information into action strength

The discussion on VS involvement in *MBL* versus *MFL* finesses the original hypothesis of the VS as a “critic,” but does little to explain the

role of the VS in regulating and invigorating motor activity (Mogenson, Jones, & Yim, 1980; Robbins & Everitt, 1996). Indeed, it would be too simple to regard the striatum as a structure receiving domain-specific information from neocortical, hippocampal, and other sources, labeling this information with reward value, and passing it on to downstream structures for further action selection. If anything in the earlier literature on VS functions stands out, it is its role in gain modulation of specific behaviors, as mediated via VS-to-VTA output, ventral pallidal-thalamic feedback loops and targets in the lateral hypothalamus, mesencephalon and pedunculo-pontine region (Groenewegen, Berendse, & Haber, 1993; Inglis & Winn, 1995; Kelley, 2004; Mogenson et al., 1980; Pennartz et al., 1994). In a Y-maze where rats used path integration to identify which of three chambers was most often rewarded (Ito et al., 2008), Lansink et al. (2012) observed that VS cells, indeed, do not simply “copy” hippocampal place-cell information and associate this with value information. Whereas CA1 neurons displayed regular place-cell mapping in this environment, VS firing did not correlate to place, but rather to action phases of the task sequence, spanning from cue light onset to goal site approach and reward consumption. Thus, the VS incorporates spatial information to encode valued actions which are appropriate to the animal's current location and context. These and other findings (Roesch, Singh, Brown, Mullins, & Schoenbaum, 2009) indicate an integration of value and motor variables in the VS.

In conclusion, the functional roles of the VS and DMS go beyond that of the “critic” in classic RL schemes. A role of the DMS in *GDB* is indicated by action–outcome devaluation studies, whereas the DLS may predominantly mediate habit formation, while not ruling out additional functions in *MBL* (cf. Stalnaker et al., 2010). Evidence on Pavlovian blocking and prospective firing activity indicates a function of the VS in *MBL* along with the DMS, raising the question in which functions the VS and DMS differ. Here we recall that neural substrates for *GDB* and *MBL* may not be identical, and whereas the DMS is implicated in *GDB*, the VS mediates motivational control and invigoration of behaviors—but not the selection of goal-directed actions per se. Conversely, coding of future reward has been found in support of *MBL* in the VS, but not DMS. Thus, the DMS and VS are functionally dissociable in multiple ways.

## 5 | DOPAMINERGIC MESENCEPHALON: REWARD PREDICTION ERROR, MODEL-BASED LEARNING, AND BEHAVIORAL REACTIVITY

### 5.1 | Introductory remarks

Schultz and colleagues famously demonstrated that ventral mesencephalic dopamine (DA) neurons in the macaque brain signal unexpected reward, as well as unexpected cues that predict subsequent reward (Mirenowicz & Schultz, 1994; Schultz, 2016; Schultz, Dayan, & Montague, 1997; Schultz, Stauffer, & Lak, 2017; Watabe-Uchida, Eshel, & Uchida, 2017). The resemblance between this phenomenology and the operation of units coding reward prediction errors in TDRL models is so striking that this algorithm is often embraced as an algorithm that

is in fact implemented by circuits involving the dopaminergic mesencephalon. Here we briefly review to what extent classic TDRL (as a particular instantiation of MFL) is generally supported by experimental evidence and whether DA signaling may also convey MBL- and GDB-related information. This is followed by a broader formulation of DA function as subserving behavioral reactivity.

## 5.2 | Dopaminergic mesencephalon: model-based versus model-free learning and goal-directed versus habitual behavior

In addition to macaque studies, recordings in rodents have tended to validate the reward prediction error hypothesis of DA neurons, as implied by TDRL models (Eshel et al., 2015; Takahashi, Langdon, Niv, & Schoenbaum, 2016) and thus support a role for DA neurons in MFL. The architecture of neural circuits feeding inputs into the VTA and processing its outputs is at least globally compatible with the requirements TDRL models pose on anatomic implementations (Berendse, Groenewegen, & Lohman, 1992; Eshel et al., 2015; Geisler, Derst, Veh, & Zahm, 2007; Menegas et al., 2015; Pennartz, 1996; Sesack & Grace, 2010; Watabe-Uchida et al., 2017). However, the “dopaminergic” model implementation of TDRL still faces a number of challenges before it can be accepted as an established fact.

First, the TDRL model assumes that dopamine would affect its presynaptic and postsynaptic targets such that it flips a molecular switch between synaptic strengthening (long-term potentiation, LTP; upon positive reward prediction errors) and weakening (long-term depression, LTD; upon negative reward prediction errors). Although several studies, mostly on dorsal striatum, have confirmed this assumption (Fisher et al., 2017; Pawlak & Kerr, 2008; Reynolds, Hyland, & Wickens, 2001; Shen, Flajolet, Greengard, & Surmeier, 2008), a multitude of other DA effects (or lack of effects) on striatal synaptic plasticity remains to be accounted for (e.g., Calabresi, Picconi, Tozzi, & Di Filippo, 2007; Hansen & Manahan-Vaughan, 2014; Pennartz, Ameerun, Groenewegen, & Lopes da Silva, 1993; Thomas, Malenka, & Bonci, 2000).

A second challenge to a dopaminergic implementation of TDRL is posed by the limited scope of reward-dependent learning behaviors that are blocked or attenuated by DA receptor antagonists (Berridge, 2007; Hagan, Alpert, Morris, & Iversen, 1983; Pennartz, 1996). Some effects on behavior, initially attributed to learning impairments, may be due to sensory, motivational, motor, and/or planning deficiencies (Denenberg, Kim, & Palmiter, 2004; Hagan et al., 1983; Pennartz, 1996; Robbins, Cador, Taylor, & Everitt, 1989). Nonetheless, dopamine signaling in VS is at least required for acquisition of conditioned reward approach, even when controlling for motor deficits (Darvas, Wunsch, Gibbs, & Palmiter, 2014; Tsai et al., 2009). Third, Redgrave, Prescott, and Gurney (1999b) noted that DA neurons may fire too early after stimulus onset to allow the subject to identify the stimulus as being reward-predictive. Recently, Schultz and coworkers (Schultz, 2016; Schultz et al., 2017) distinguished an early and late dopaminergic component in response to stimuli, reflecting the physical impact of stimulus detection and value prediction error, respectively. The early

component raises two interesting issues on its function and consequences. First, because this component is value-independent, it is more compatible with a role of DA cells in early behavioral reactions to salient environmental changes, prompting, for example, saccades toward the object. Second, the physical impact of any salient stimulus should raise dopamine release via the early component and, according to TDRL, would thereby induce a synaptic modification as if a “positive surprise” signal had been present, with potentially dysfunctional consequences.

Coming back to the MBL–MFL distinction, the above findings do not rule out a role for dopamine neurons in MBL. Indeed, they do not only signal error in value prediction but also errors in the prediction of sensory features of expected reward (Takahashi et al., 2017; cf. Bromberg-Martin, Matsumoto, & Hikosaka, 2010). Assuming that these sensory prediction errors can guide adaptive behavior, this places the DA system in the domain of both MBL and MFL.

As regards GDB versus HB, different DA functions have been studied in relation to distinct target areas of the mesencephalic dopaminergic projections (in view of the fact that complete loss of DA function leads to severe motor incapacitation and starvation; Darvas et al., 2014). These studies reveal that DA can support both GDB and habit formation in a way that co-depends on the function of the target area in each type of behavior. For instance, bilateral 6-hydroxydopamine (6-OHDA) lesions of the nigrostriatal pathway mainly targeting DLS maintained sensitivity to reward devaluation, indicating a function in habit formation (Faure, Haberland, Conde, & El Massioui, 2005; consistent with evidence that potentiating DA release by amphetamine accelerates habit formation; Nordquist et al., 2007). In contrast, pretraining 6-OHDA lesions of prelimbic (but not infralimbic) cortex caused a deficit in adapting instrumental responses to changes in action–outcome contingency (Naneix, Marchand, Di Scala, Pape, & Coutureau, 2009), consistent with the role of prelimbic cortex in GDB. The same study also reported that instrumental responses remained sensitive to outcome devaluation under the same treatment, showing a dissociation between two hallmarks of GDB and thus suggesting that GDB is not mediated by a unitary mechanism. Yet a different study using DA receptor stimulation found a contrasting result: Whereas infralimbic infusions of DA amplified goal-directed responding in an outcome-devaluation paradigm, prelimbic manipulation had no such effect (Hitchcott, Quinn, & Taylor, 2007). Whether the differences with the Naneix et al. study are attributable to the overall balance of DA receptor functions affected by DA infusion, to differences in devaluation procedures or other factors, remains to be investigated.

## 5.3 | Further analysis: Dopamine and behavioral reactivity

In addition to these unresolved questions, another enigma still remains: The relationship between DA neurons as reward prediction error coding units vis-à-vis the well-known role of dopamine in movement initiation, posture regulation and other aspects of motivated motor behavior, as affected in Parkinson's disease. This seemingly

dual role has been explained by the hypothesis that healthy motor behaviors are maintained by low-level tonic DA neuron firing activity, whereas learning effects would be mediated by burst activity, resulting in strong DA release (Schultz, 2016). However, phasic dopamine neuron firing has also been associated with motor action, at least in a general sense (Schultz et al., 2017). This activation is associated with global limb and head movements or with small-scale movements such as licking and chewing (DeLong, Crutcher, & Georgopoulos, 1983; Schultz, Ruffieux, & Aebischer, 1983). Movement-related changes in DA cell firing have been somewhat ignored recently because of an apparent lack of consistency, and their predominant absence during simpler tasks such as Pavlovian conditioning, but spontaneity in complex behaviors may be a significant aspect of DA function, compromised as it is in Parkinson's disease. While many aspects of DA function in the MBL–MFL and GDB–HB distinctions await further testing, another critical question thus remains, namely which properties of graded DA-release mechanisms determine the putative boundary between motor-related versus reward prediction error related effects on DA release, and whether in fact any “hard” boundary can be delineated.

An alternative hypothesis holds that the basic function of dopamine is to enable behavioral reactivity to salient, unexpected sensory input in general—visual, auditory, proprioceptive, or otherwise—which aligns better with the motor deficits observed in Parkinson's disease (cf. Pennartz, 1996; Redgrave et al., 1999b; Robbins & Everitt, 1982; Rodriguez-Oroz et al., 2009; Salamone, Cousins, & Bucher, 1994). As in the reward prediction error hypothesis, this account holds that DA neurons signal prediction errors, but these are of a more generalized nature, as they include both motivational (value-related) errors and errors in sensorimotor predictions. The underlying rationale is that, functionally, surprising sensory changes require further exploratory, proactive and reactive movements, such as saccades, grabbing movements, locomotion, and postural adjustments.

Evidence for the behavioral reactivity hypothesis of DA comes, first, from the “early” dopamine response component and the movement-related DA firing responses already mentioned, and, second, from studies reporting that laterally located mesencephalic DA neurons, mostly in the Substantia Nigra pars compacta (SNPC), are sensitive to salient and unexpected, but *neutral sensory* stimuli, whereas VTA cells respond to unexpected *reward* (Bromberg-Martin et al., 2010; Pennartz, Ito, et al., 2011). Indeed, the SNPC receives predominantly excitatory inputs from the somatosensory and motor cortices, whereas the VTA is heavily innervated by the lateral hypothalamus (Watabe-Uchida, Zhu, Ogawa, Vamanrao, & Uchida, 2012). Third, studies monitoring extracellular DA levels in striatum using fast-scan cyclic voltammetry have noted marked deviations in DA signaling from transient, reward prediction error related firing of DA cells, emphasizing correlates and causal functions in the biasing of action selection (Howard, Li, Geddes, & Jin, 2017), reward expectancy, response invigoration, and the estimation of value versus costs of actions and internal operations (Berke, 2018).

The behavioral reactivity hypothesis is somewhat akin to the incentive–salience hypothesis (Berridge, 2007) in its emphasis on the

motivational (“wanting”) function of the DA system, although this hypothesis holds that dopaminergic mechanisms attribute incentive salience specifically to reward-related stimuli, not to salient or unexpected stimuli in general. The behavioral reactivity account seamlessly matches another critical point touching upon the scope of dopamine in general brain function: The expression of reward prediction error signaling by DA cells may just be the tip of an iceberg. The processing of unexpected reward- and sensory-related signals is so essential for the survival and reproduction of animals that a wealth of brain areas is equipped with mechanisms to react to unexpected cues, contexts, movement, and outcomes, interdigitating with the specialized functions of each area (cf. Pennartz, 1997). For instance, neurons in layer II–III of mouse visual cortex code sensory prediction errors (Keller, Bonhoeffer, & Hubener, 2012; cf. Bastos et al., 2012; Leinweber, Ward, Sobczak, Attinger, & Keller, 2017). Furthermore, brain-wide fMRI studies suggest that reinforcement-related signals may be ubiquitous throughout the cortex (Serences, 2008; Vickery, Chun, & Lee, 2011). Even neurons in primary sensory cortex show reward-expectancy correlates and reward-dependent learning effects on sensory tuning and retinotopic mapping (Goltstein, Coffey, Roelfsema, & Pennartz, 2013; Goltstein, Meijer, & Pennartz, 2018; Shuler & Bear, 2006; cf. Bao, Chan, & Merzenich, 2001). Thus, coding of prediction errors may be so ubiquitous across the brain that an exclusive attribution of this function to DA cells might be the result of “searching under the streetlight.”

In conclusion, the original evidence on reward prediction error coding by DA neurons in the VTA remains firmly standing, which however does not imply that the mesolimbic DA system therefore exclusively functions to mediate MFL through TDRL. Evidence for sensory-specific coding, area-specific dopamine effects on GDB and motivational correlates of extracellularly recorded dopamine levels suggest a broader role of DA neurons, pointing to a more general functional repertoire subserving MBL, goal-directed actions and the overarching concept of behavioral reactivity.

## 6 | HIERARCHICAL CONTROL OF BEHAVIOR THROUGH TOPOGRAPHICALLY ORGANIZED CORTICO-BASAL GANGLIA-THALAMIC LOOPS

### 6.1 | Introductory remarks

In this section, we will discuss in more detail how the brain areas, individually reviewed above, interact to accomplish hierarchically organized GDB. How do brain systems for global action policies and long-term planning of sequential behavior control subroutines, carried out as short-lasting, elemental sensorimotor skills? (Barto & Mahadevan, 2003; Botvinick, 2008; Dezfouli, Lingawi, & Balleine, 2014; Pezzulo et al., 2014). The evidence reviewed so far is consistent, first, with the engagement of HPC, PFC, and the ventromedial striatal region in behaviors requiring MBL (while not excluding MFL), whereas the DLS is more clearly linked to MFL (not excluding MBL); dopamine neurons may rely on both MBL and MFL. Second, the evidence suggests that

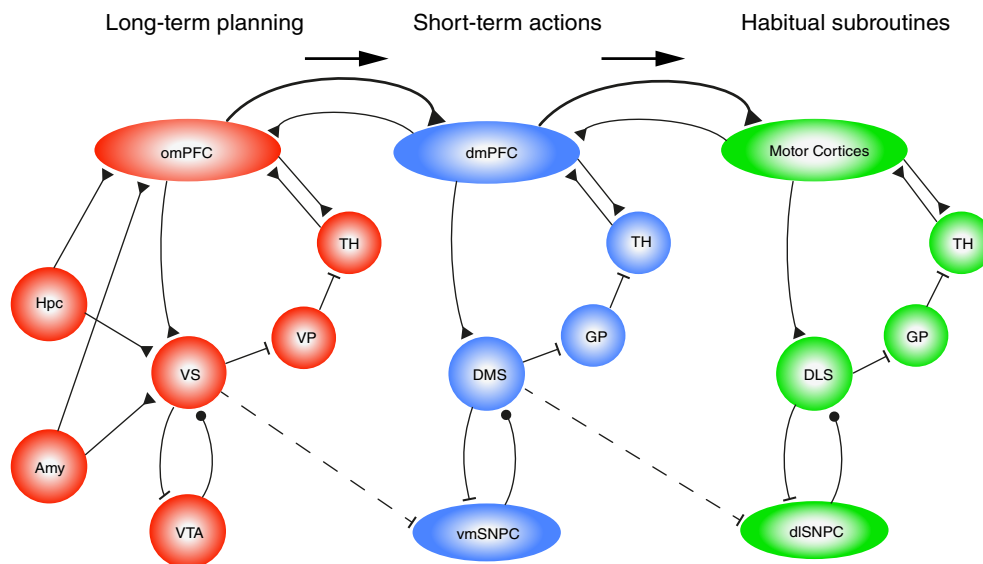
GDB and MBL—despite their differences in provenance and conceptualization—share a significant number of common neural substrates, at least when defined at the coarse level of structures or regions. Therefore, we will regularly refer to neural substrates mediating both GDB and MBL in conjunction below.

## 6.2 | Hierarchical organization of cortico-basal ganglia loops

Previous proposals on hierarchically organized behavior mostly focused on the PFC and particularly on its dorsolateral regions. In particular, a topographic organization was distinguished within the frontal cortex, with higher levels of behavioral control being associated with rostral PFC areas and lower levels to caudal regions (Azuar et al., 2014; Botvinick, 2008; Koehlin & Hyafil, 2007; but see Badre & Nee, 2018). Adhering to the concept of hierarchical RL (Barto & Mahadevan, 2003; Botvinick, 2008; O'Reilly & Frank, 2006), low-level behaviors or subroutines are temporally organized by nesting them in higher-level representations of more global behaviors and this process would be mediated by the PFC (Botvinick, 2008; Botvinick, Niv, & Barto, 2009). Here, we emphasize that hierarchical behavioral control includes more than the (hierarchical organization of) classic RL alone,

tied as this is to MFL. The question arises: How can GDB and MBL be fit in?

While in hierarchical RL top-down control may be implemented by a rostrocaudal direction of connectivity in the PFC, it is less clear how the basal ganglia, HPC and associated structures such as the amygdala can be incorporated, and how sensorimotor subroutines are integrated into hierarchically organized behavior. Both in primates and rodents multiple cortico-basal ganglia-thalamic loops have been identified (Alexander, Crutcher, & DeLong, 1990; Groenewegen, Berendse, Wolters, & Lohman, 1990; Groenewegen & Uylings, 2000; Voorn et al., 2004). Classically, the loops distinguished in primates are (a) a “limbic”-affective loop including the anterior cingulate cortex and medial OFC; (b) dorsolateral and lateral orbitofrontal loops subserving cognitive functions such as working memory and attentional control; (c) an oculomotor loop comprising the frontal eye field and supplementary eye field; and (d) a motor loop comprising the motor cortex, supplementary motor area and premotor cortex (Alexander et al., 1990). In rodents, a similar distinction in loops is made, with (a) a limbic-affective loop that originates primarily in orbitofrontal and ventral medial prefrontal areas (here abbreviated as omPFC), which mainly project to the VS (core and shell), (b) a more exteroceptively and cognitively oriented circuit originating in dorsal-medial prefrontal



**FIGURE 2** Interactions between neural systems for (a) long-term goal-setting and planning, (b) short-term actions, and (c) executing habitual subroutines. The leftmost system (in red) has orbitofrontal and ventromedial prefrontal structures (omPFC) and ventral striatum (VS) at its core, forming a limbic-affective corticobasal ganglia loop that is proposed to mediate long-term goal setting and planning to obtain outcomes desired on the long term. This loop is supported by episodic memory information retrieved via the hippocampus (Hpc) and positively or negatively valued information from the amygdaloid complex (Amy). OmPFC, dmPFC, and motor cortices are reciprocally connected, yet in terms of hierarchical control, it is proposed that omPFC exerts a top-down control over dmPFC, which in turn controls the motor cortices (symbolized by arrows and stronger projections going rightward in the scheme). The dmPFC forms a loop (in blue) with the dorsomedial striatum (DMS), whereas the motor cortices form more caudally located loops (in green) with the dorsolateral striatum (DLS; convergence from sensory cortices onto DLS is not shown here). The hierarchical control from omPFC to dmPFC and motor cortical loops is reinforced by inhibitory outputs from striatal structures to parts of the dopaminergic midbrain specifically involved in these respective loops (VS inhibits ventromedial substantia nigra pars compacta, vmSNPC; DMS inhibits the dorsolateral substantia nigra pars compacta, diSNPC). Excitatory, glutamatergic connections are represented by black triangular terminals; inhibitory GABAergic connections by flat endings; modulatory dopaminergic projections by black circular terminals. Note that this scheme primarily follows rodent brain organization, but that it can be applied to primates with some modifications. VP, ventral pallidum; GP, globus pallidus; TH, thalamus; VTA, ventral tegmental area

areas (dmPFC; mainly dorsal prelimbic cortex, anterior cingulate cortex and area Fr2), which project to the DMS; and (c) a motor loop involving sensorimotor cortical areas projecting to the DLS (Flaherty & Graybiel, 1995; Groenewegen et al., 1990; Voorn et al., 2004). While these loops can be subdivided into finer sub-loops, the most relevant partition here is that of the rodent PFC into omPFC and dmPFC (Groenewegen & Uylings, 2000; Voorn et al., 2004).

We propose that the limbic-affective loop, including omPFC-VS circuits, occupies the highest position in a hierarchy of loops (Figure 2). The next highest level is the loop comprising dmPFC and DMS, which mediates cognitive operations (e.g., working memory and attentional set-shifting) and GDBs on the short term. In defining these loops, we primarily follow rodent prefrontal organization, noting that rat dmPFC bears functional similarities to primate dorsolateral PFC (Uylings, Groenewegen, & Kolb, 2003) but also shares anatomic features with primate orbitomedial PFC (Heilbronner, Rodriguez-Romaguera, Quirk, Groenewegen, & Haber, 2016; Preuss, 1995; Wise, 2008). In humans, the frontopolar cortex may contribute to omPFC rather than to dmPFC-like circuits (Gläscher et al., 2012).

These two high-level loops may not only control lower level, sensorimotor loops (Figure 2, rightmost module) through direct cortico-cortical top-down connections but also via selection mechanisms in the basal ganglia. These mechanisms may comprise lateral (or recurrent) inhibition between striatal medium-sized spiny neurons (Burke, Rotstein, & Alvarez, 2017; Plenz, 2003; Taverna, van Dongen, Groenewegen, & Pennartz, 2004; van Dongen et al., 2005) and other, interneuron-dependent inhibitory operations in the striatal-pallidal “funnel” (Bar-Gad, Morris, & Bergman, 2003; Taverna, Canciani, & Pennartz, 2007). With “funnel” we mean that, when descending along the cortical-striatal-pallidal stages of each loop, the cell count dramatically decreases. With this reduction comes an increased degree of convergence of anatomical projections onto a small pallidal volume (Bar-Gad et al., 2003; Pennartz et al., 1994). Thus, between-loop interactions may also occur at the level of the globus pallidus and its interactions with the subthalamic nucleus and substantia nigra pars reticulata (Bugaysen, Bar-Gad, & Korngreen, 2013; Sadek, Magill, & Bolam, 2007; Sato, Lavallee, Levesque, & Parent, 2000), or between thalamic subregions receiving basal ganglia outputs.

The outputs from the limbic-affective loop are proposed to steer processing in more caudal and dorsal loops, viz. the dmPFC-to-DMS loop and the sensorimotor cortices-to-DLS loop. This results in short-term GDB and habitual subroutines being controlled by higher-level mechanisms for long-term goal setting and planning. The DLS conforms to this layout, even though its topographic location in the striatum is not strictly “caudal.”

Placing the limbic-affective (omPFC) loop at a higher level of control than the more “cognitive”, exteroceptive and action-outcome-oriented (dmPFC) loop may seem surprising, but is motivated by evidence that omPFC areas are heavily involved in achieving the organism's long-term goals, associated as these are with homeostatic variables and basic motivational drives (i.e., to satisfy hunger, thirst, sex, to avoid pain etc.; Carmichael & Price, 1996; Critchley & Rolls, 1996; Groenewegen & Uylings, 2000; Pennartz et al., 1994). In a

strong functional-evolutionary sense, cognitive operations such as working memory, short-term action choices and attention are subordinate to achieving long-term motivational goals. This high-level position is further supported by evidence on ventral and mPFC lesions in humans, pointing to dysregulation of value-based decision-making in general (Gläscher et al., 2012) and on the key role of omPFC-VS circuits in balancing behavioral policies on long-term versus short-term time scales (borne out by patterns of impulsivity and preference for delayed reward; Cardinal, Pennicott, Sugathapala, Robbins, & Everitt, 2001; Jimura, Chushak, & Braver, 2013; Mar, Walker, Theobald, Eagle, & Robbins, 2011). Moreover, omPFC areas have strong connections with the hypothalamus and various brain stem centers implicated in regulation of basic homeostasis and autonomous functions (Carmichael & Price, 1996; Groenewegen & Uylings, 2000). This proposal aligns well with the role of omPFC and VS in MBL and GDB. Thus, there exists no tension between the emotional connotations ascribed to orbitomedial prefrontal structures and their having a position high in the behavioral control hierarchy—on the contrary. Having said this, it should be emphasized that reciprocal control relations between omPFC and dmPFC likely exist. Further arguments for proposing this hierarchical arrangement are given below.

### 6.3 | Hippocampal, amygdala, and dopaminergic outputs to cortico-basal ganglia loops

How may the HPC, amygdala and dopaminergic mesencephalon fit into this proposal? With its capacities for mapping world-state variables and MBL, the hippocampal formation sends output to the omPFC and VS, whereas hippocampal-subicular output to the DLS and its associated sensorimotor loop is much scarcer (Groenewegen et al., 1987). This way, task-space information in PFC is enriched with hippocampal information on world states and relationships between state variables, helping to identify expected outcomes and select which task rules and goals apply to the agent's current environmental context during planning and execution of GDB (see Wikenheiser, Marrero-Garcia, & Schoenbaum, 2017, for a causal influence of ventral subiculum on OFC coding of expected outcome). This proposal may provide a solution for the previously raised observation that the HPC is involved in MBL but not required for GDB per se (see section on Hippocampus): By supplying the PFC with world state information acquired through MBL, the HPC may facilitate prefrontal mechanisms for implementing GDB without being a neural substrate necessarily and causally required for GDB itself (as defined in the Introduction). That Corbit et al. (2002) failed to observe any effect of excitotoxic hippocampal lesions on instrumental performance may thus be explained by assuming that world-state knowledge provided by the HPC is not causally required to solve their particular task (viz. pressing two levers which were each coupled to delivery of a unique food outcome, followed by procedures for testing outcome devaluation and degradation of instrumental contingency).

Vice versa, the PFC sends signals to the medial temporal lobe, including HPC and parahippocampal regions, which query the stored database on world states, and stimulate memory retrieval and internal

simulations of potential future scenarios, as expressed by replay and theta look-ahead sequences (Pezzulo et al., 2014; Redish, 2016). How this query-and-retrieval process is implemented is unknown, although PFC output has been shown to influence hippocampal spatial information processing, as indicated by optogenetic–electrophysiological studies (Ito et al., 2015; Schmidt et al., 2019) and PFC lesion effects on area CA1 place field stability (Kyd & Bilkey, 2003).

Hippocampal outputs to the omPFC-VS loop may also bias the initiation and invigoration of goal-directed actions by selection mechanisms in the basal ganglia (Ito et al., 2008; Lansink et al., 2012; Robbins & Everitt, 1996). Similarly, amygdala output will affect information processing in this loop, but also in the more short-term, cognitively oriented dmPFC-DMS loop, for example, by conveying information on stimulus value to mechanisms for instrumental action selection (Cardinal et al., 2002; Hall et al., 2001).

For dopaminergic outputs, we propose that the VTA—with its dense outputs to the high-level omPFC-to-VS loop—signals “affective surprise” (error in reward prediction, but enriched with MBL components), which conforms to the motivational-affective nature of this loop. In contrast, DA cells in SNPC—projecting to the dmPFC-to-DMS and sensorimotor cortical-DLS loops—are suggested to signal “sensorimotor surprise,” i.e., errors in the prediction of sensorimotor states resulting from the agent's actions or from external events. The SNPC output to the sensorimotor-DLS loop thereby conforms to MFL and HB. Both kinds of surprise may subserve learning and behavioral reactivity to enable agents to quickly adapt their behavior and posture once relevant and unpredicted environmental changes occur. This dual function also holds for multi-step tasks in which, during progressive learning, reward-predictive (CS+ related) and sensorimotor-predictive events come to function as surrogate prediction errors.

The notion of behavioral reactivity entails that patterns of ongoing, routine behavior can be interrupted by motivationally relevant stimuli to allow adaptive changes in behavior. Our hypothesis on hierarchical behavior proposes that switching in behavioral patterns due to unexpected changes in reward prediction is accompanied by a rebalancing of mesencephalic DA release across its diverse target areas, such that novel, GDB is invigorated at the expense of habits (which, however, will be facilitated again when the subject falls back on routine behaviors). Similarly, sensorimotor surprise signals will induce adjustments in subroutines in order to achieve the low-level goal of an HB. Interactions between DA release patterns and the activity in the direct and indirect striatal pathways during behavioral switching (cf. Nonomura et al., 2018) remain to be investigated.

The integration of DA function in the scheme of Figure 2 offers further means to address how high-level loops control lower-level loops. In addition to the mechanisms already mentioned, the VTA in rats—with its predominant inputs from omPFC and VS—projects not only back to these structures, but also to the DLS (Maurin, Banrezes, Menetrey, Mailly, & Deniau, 1999; Pennartz et al., 2009). Moreover, the “ascending spiral” from VS-VTA levels up to the lateral SNPC, projecting to DLS, is known well from primate studies (Haber, Fudge, & McFarland, 2000). This asymmetric organization offers yet another

argument to place the limbic-affective loop at a higher hierarchical position than the dmPFC-to-DMS and sensorimotor cortical-to-DLS loops. This aligns with the adaptive changes in the striatum observed in drug addiction, progressing from medial to dorsolateral striatal sectors (Belin-Rauscent, Everitt, & Belin, 2012).

In sum, we propose that hierarchical behavioral control arises from top-down control by the MBL- and GDB-based omPFC-VS loop over the dmPFC-DMS loop, which in turn controls the sensorimotor cortical-DLS loop involved in HB. Conversely, information on low-level routines may be transmitted to high-level planning systems (Pezzulo et al., 2014) via corticocortical or intrastriatal interactions (Figure 2). As compared to previous proposals, the current hypothesis has the advantage of utilizing known cortico-basal ganglia loops in implementing top-down control.

## 7 | COMMUNICATION BETWEEN PREFRONTAL CORTEX, HIPPOCAMPUS AND VENTRAL STRIATUM DURING ENCODING AND OFF-LINE MEMORY CONSOLIDATION

### 7.1 | Introductory remarks

A key requisite for a multi-area network to control hierarchically organized behavior is to have appropriate communication mechanisms in place during memory encoding and consolidation. Here, we will limit the discussion to a brief overview of interactions between some of the key players reviewed so far: HPC, PFC, and VS. These structures function as hubs communicating with each other and with their wider distributed network in different neurophysiological modes characterized by rhythmic oscillations, detected via coherent LFP and spiking activity (Benchenane et al., 2010; Fujisawa & Buzsáki, 2011; Jones & Wilson, 2005; Lansink et al., 2016; van Wingerden et al., 2010a; van Wingerden, Vinck, Lankelma, & Pennartz, 2010b; Young & Shapiro, 2011).

### 7.2 | Communication during active behavior

During ongoing behavior, interactions between the *HPC*, *medial prefrontal*, and *orbitofrontal cortex* are particularly manifest by synchrony in the theta range (6–12 Hz in rodents). For instance, Jones and Wilson (2005) showed that mPFC neurons fire in synchrony with theta oscillations recorded from dorsal area CA1, and this spike-LFP synchrony is especially strong during behavior taxing spatial working memory (O'Neill, Gordon, & Sigurdsson, 2013). Hippocampal–mPFC theta synchrony may be instrumental during both encoding and retrieval phases of memory processing in the awake state. Theta rhythm may function to dynamically open and close communication channels between these two areas (Benchenane et al., 2010) and theta phase precession (Huxter, Burgess, & O'Keefe, 2003) provides a mechanism to feed world-state representations sequentially into mPFC.

However, one should be cautious in applying a simplified concept to hippocampal–cortical communication as if “memory transfer”

would occur from one source to the other. We deem this unlikely because hippocampal input to the mPFC probably constitutes only a minor fraction of the total synaptic inputs received by PFC pyramidal neurons, such that their spike output is determined by the integration of a large, mixed plethora of inputs, also of nonhippocampal origin (e.g., from amygdala, thalamus, contralateral mPFC, OFC, and sensorimotor cortices; Carmichael & Price, 1996; Dembrow, Zemelman, & Johnston, 2015; Gabbott et al., 2012; Groenewegen & Uylings, 2000; Hoover & Vertes, 2007). Recalling the PFC functions discussed above, we rather propose that hippocampal information on world states is integrated with sensory, motor, goal and value-related inputs, arising from nonhippocampal sources, to enable the PFC to encode the agent's current task space and to help compute optimal plans and decisions. In other words, world-state information is insufficient to code the materials on which decisions are based; egocentric sensorimotor, motivational, and other information are needed to define action policies.

During active behavior, theta rhythmicity is also characteristic of *hippocampal-ventral striatal* communication, with subsets of ventral striatal cells firing coherently with hippocampal theta rhythm. Both phase-locking and phase-precession of VS neurons to hippocampal theta rhythm have been reported (Berke, Okatan, Skurski, & Eichenbaum, 2004; Lansink et al., 2016; Lansink, Goltstein, Lankelma, McNaughton, & Pennartz, 2009; van der Meer & Redish, 2011a). Additional beta-synchronized activity (15–25 Hz) recorded from both area CA1 and VS correlated with stimulus-triggered approach of goal sites and is interpreted as an intensified mode of HPC-VS communication (Lansink et al., 2016). Next to theta and beta rhythmic activity, lower (~4 Hz) frequency bands have been reported to coordinate activity in PFC–HPC–VTA circuits (Fujisawa & Buzsáki, 2011). When comparing oscillatory phenomena between the HPC–PFC–VS system and the sensorimotor–DLS loops, the organizing role of theta rhythm reported for the HPC–PFC–VS system appears to be largely lacking in the sensorimotor–DLS system (Berke et al., 2004; Lalla, Rueda Orozco, Jurado-Parras, Brovelli, & Robbe, 2017), which correlates with the segregation in an MBL/GDB-based and MFL/HB system.

### 7.3 | Off-line processing and the standard model of memory consolidation

Next, we pose a similar question as in the discussion on hierarchically organized behavior: Can we discern a hierarchy or at least directionality in brain structures organizing off-line memory processing? With “off-line processing,” we denote episodes where active task performance is absent: Quiet wakefulness, pauses interleaved with active behavior, and sleep (focusing on non-REM sleep in the current context).

Current conceptualizations of memory consolidation often refer to the “standard consolidation model” (Frankland & Bontempi, 2005; cf. Battaglia, Benchenane, Sirota, Pennartz, & Wiener, 2011; Buzsáki, 1989; Marr, 1971), holding that long-term declarative memory is only transiently dependent on HPC. Newly acquired episodic memories would be temporarily stored in HPC and gradually transferred to

cortical regions during consolidation. Apart from the problem of synaptic matrix integration, signaled above, this account has left unresolved a number of difficulties, one of which is the question whether and how memory traces would vanish from the HPC. The alternative scenario holds that this structure retains the original traces, whereas it is the subject's behavior that becomes less dependent on hippocampal integrity, as transformations (e.g., generalization, semanticization, automatization) in HPC-receptive structures enable subjects to behave in a non-HPC dependent manner (Battaglia & Pennartz, 2011; Moscovitch et al., 2005; Tse et al., 2007; Winocur, Moscovitch, & Bontempi, 2010). Furthermore, recent evidence suggests that engrams are formed in the HPC and neocortex in a temporally overlapping manner (Kitamura et al., 2017).

Regardless of the validity of the standard consolidation model, replay of task-related activity patterns has been a critical subject in investigating memory retrieval and consolidation. Sharp-wave ripples (SWRs) in HPC have been suggested to play a network-synchronizing role during off-line consolidation (Girardeau, Benchenane, Wiener, Buzsáki, & Zugaro, 2009; Johnson et al., 2012; Kudrimoti, Barnes, & McNaughton, 1999; Lansink et al., 2009; Pennartz et al., 2004; Tamminen, Lambon Ralph, & Lewis, 2013), particularly during non-REM sleep and behavioral pausing. Hippocampal ripples are waxing and waning high-frequency oscillations (150–200 Hz) occurring during immobility, consummatory behaviors and non-REM sleep (Buzsáki, 1986; O'Keefe & Nadel, 1978). During post-task non-REM sleep, hippocampal reactivation of place-cell patterns was enhanced during SWRs as compared to nonripple intervals (Kudrimoti et al., 1999; Wilson & McNaughton, 1994). Reactivation of neural patterns during post-task non-REM sleep has not only been observed in HPC, but also in visual cortex, amygdala, VTA, VS, mPFC, OFC, and parietal cortices—often in conjunction with hippocampal reactivation (Euston et al., 2007; Girardeau, Inema, & Buzsáki, 2017; Gomperts, Kloosterman, & Wilson, 2015; Ji & Wilson, 2007; Lansink et al., 2008; Lansink et al., 2009; Pennartz et al., 2004; Qin, McNaughton, Skaggs, & Barnes, 1997; Rusu et al., 2016; Tang, Shin, Frank, & Jadhav, 2017; Valdes, McNaughton, & Fellous, 2015). Ripple-associated replay has been suggested to subserve consolidation of memories pertaining to reward and goal-directed trajectories (Foster & Wilson, 2006; Kalenscher & Pennartz, 2008). Causal intervention experiments have underscored the impact of ripple-associated firing on memory operations during sleep or task performance (Ego-Stengel & Wilson, 2010; Girardeau et al., 2009; Jadhav et al., 2012).

Papale, Zielinski, Frank, Jadhav, and Redish (2016) showed that ripple density was inversely correlated with deliberative, VTE-type of behavior. This suggests that the cognitive processes underlying model-based behavior, such as prospective simulation of future trajectories (Pezzulo et al., 2014) cohere with hippocampal theta-look ahead sequences rather than with fast, SWR-associated replay. However, this does not imply that SWR-associated replay would have no relationship with GDB and MBL. Because hippocampal replay pertains to chains of sensory-specific states, it is consistent with MBL, yet it is too early to conclude that it would therefore not involve MFL. Indeed,

reactivation and memory consolidation may be sustained by many brain structures independently of the HPC. For instance, memory consolidation in the somatosensory motor network, associated with MFL, can occur without hippocampal involvement (Miyamoto et al., 2016).

#### 7.4 | Off-line processing: From hippocampal to corticothalamic control

Despite the importance of SWRs for mnemonic operations, the interactions between HPC and its target areas cannot be characterized as one-way traffic, because intra-hippocampal activity is also controlled by inputs arising from within its larger network, including the neocortex. For instance, optogenetic burst-stimulation in the VTA of mice exploring a novel environment was shown to enhance reactivation of hippocampal ensembles during post-task sleep and rest (McNamara, Tejero-Cantero, Trouche, Campo-Urriza, & Dupret, 2014), suggesting that dopamine cell discharge promotes subsequent sleep reactivation. A second example is provided by the bidirectionality of neocortical-hippocampal interactions during reactivation (Sirota, Csicsvari, Buhl, & Buzsaki, 2003; cf. Battaglia et al., 2011; Rothschild, Eban, & Frank, 2017).

However, even causally intervening experiments leave space for a scenario where the physiological cascade leading to memory consolidation does not originate in the HPC, but is initiated in thalamocortical circuits where spindles—embedded in Up states—organize hippocampal ripple firing in time (Latchoumane, Ngo, Born, & Shin, 2017). Moreover, we still lack strong causal evidence that the HPC regulates replay in target structures such as PFC and VS. Thus, we argue that the “hippocampocentric” view of off-line replay should be broadened to include large-scale activity in the corticothalamic network in order to better understand how multi-area network activity underlying MBL and GDB is organized. Hippocampal activity strongly depends on inputs from the thalamocortical network where, during slow-wave sleep, oscillatory activity is highly synchronized (Crunelli & Hughes, 2010; Huguenard & McCormick, 2007). Indeed, hippocampal SWRs and associated PFC replay correlate with the timing of cortical slow waves and down-to-up state transitions (Battaglia, Benchenane, Sirota, Pennartz, & Wiener, 2011; Peyrache et al., 2009).

In humans, 0.75 Hz transcranial stimulation promotes grouping of slow spindle activity (8–12 Hz) during Up states and sequential organization of SWRs at spindle troughs (Marshall, Helgadóttir, Mölle, & Born, 2006; Mollé & Born, 2011). In rats that were trained on an object-in-place task, Maingret, Girardeau, Todorova, Goutierre, & Zugaro (2016) used electrical stimulation of motor cortex to boost synchronized SWR, delta-wave and spindle activity. This treatment enhanced memory consolidation and prefrontal neural responses. Moreover, when spindles were induced using optogenetic stimulation of the thalamic reticular nucleus, memory consolidation for contextual (but not cued) fear conditioning was improved. This effect was observed only when stimulation coincided with cortical, slow-oscillation Up states whereas spindle inhibition decreased memory performance (Latchoumane et al., 2017). Further suggestions for a causal role of sleep oscillations in consolidation were obtained by slow

oscillatory, transcranial direct-current stimulation in humans with mild cognitive impairment (Ladenbauer et al., 2017). Thus, evidence is accumulating to suggest a temporally organizing role of spindle and slow-wave activity in memory consolidation.

#### 7.5 | Frontal-to-caudal organization of memory processing during slow-wave sleep

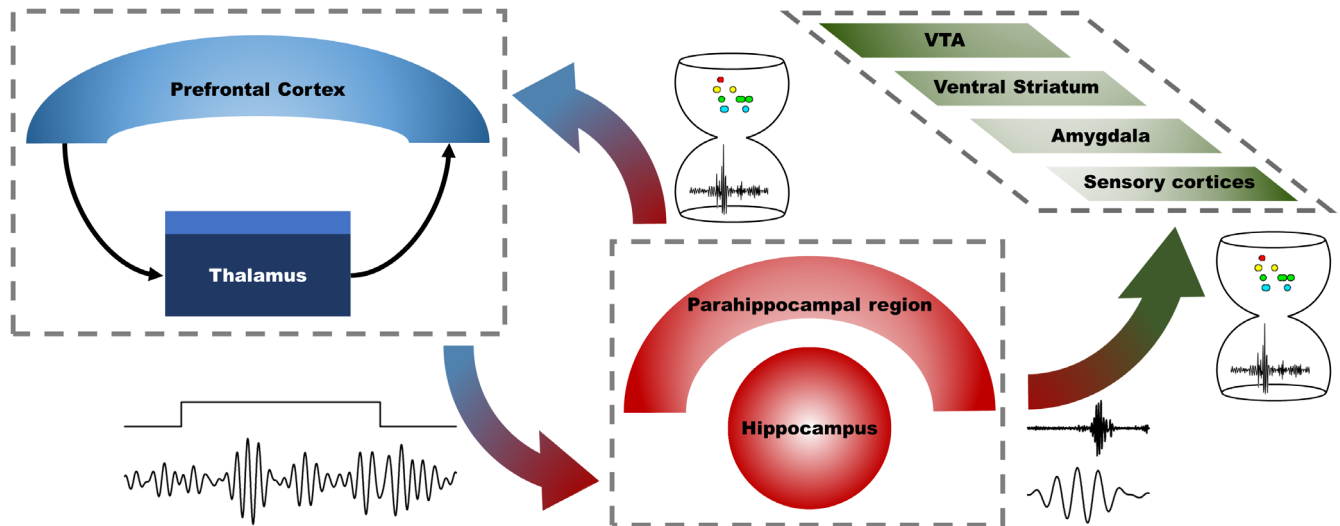
In humans, slow waves have been reported to travel at a speed of 1.2–7.0 m/s mainly from prefrontal-orbitofrontal to posterior neocortical areas (Massimini, Huber, Ferrarelli, Hill, & Tononi, 2004). In EEG and single-unit recordings from epileptic patients, slow waves were manifested locally, but also propagated from mPFC to temporal lobe areas, including HPC (Nir et al., 2011). In addition, they entrained the thalamus, favoring spindle development (Luthi, 2014). When spindles were generated independently of Up states in isolated cortical preparations, slow oscillations promoted a temporally coherent organization of spindle volleys across widespread neocortical regions (Luthi, 2014).

Based on this admittedly restricted evidence, we propose two functional loops regulating slow-wave sleep consolidation. The first is an overall “initiation loop” activated by slow waves traveling across the neocortex which, in association with thalamic spindles, leads to selective recruitment of neocortical ensembles in the Up state, which feed outputs into the (para)hippocampal network. Importantly, waves are propagated in a frontal-to-caudal direction, which matches the direction we propose for hierarchically organized activity in cortico-basal ganglia-thalamic loops (Figure 2). By this directionality, prefrontal ensembles concerned with the encoding of task space will be recruited first, followed by recruitment of more caudally located ensembles, subordinate in the behavioral hierarchy and more involved in short-term behaviors and low-level subroutines.

Second, we propose that a nested, thalamocortically entrained “hippocampal loop” reactivates stored information using SWRs and emits the resultant activity to target areas (Figure 3). By acting in concert with neocortical areas and recruiting other modules (e.g., PFC, VS, VTA, nucleus reuniens, and amygdala) the HPC may thus orchestrate memory trace consolidation in its directly or indirectly connected network. Lesion and pharmacogenetic inactivation experiments allow us to finesse the role of HPC in consolidation. Using an odor-sequence recognition task, Fortin et al. (2002) showed that memory for sequences was impaired following hippocampal lesions, whereas single-odor recognition performance remained intact. Recently Barker et al. (2017) applied pharmacogenetic inactivation of CA1-to-mPFC projections to an object-location task including a temporal component, and found that the dorsal and intermediate CA1-mPFC projections differentially contribute to temporal order judgment and spatial memory, respectively. These results align with an electrophysiological study on mice with a CA1 NMDA receptor knockout (see above, Cabral et al., 2014) reporting that these animals are selectively impaired on memory for long, but not short, behavioral sequences, with concomitant place-cell mapping deficiencies.

Thus, instead of corroborating the HPC-to-neocortical “memory transfer” hypothesis, these data support the notion of the HPC as a





**FIGURE 3** Coordination of sleep replay in hippocampus and associated structures by activity in cortico-thalamic loops. On the left-hand side, spindle activity is proposed to be coordinated by prefrontal-thalamic loops. The dark blue part of the thalamus represents relay nuclei, lighter blue is the reticular nucleus of the thalamus. This spindle activity (“initiation loop”; represented by LFP trace at the bottom left, recorded from a tetrode placed in orbitofrontal cortex, filtered at 10–18 Hz; total duration: 2.0 s) controls inputs to the hippocampus (and parahippocampal regions such as ento- and perirhinal cortex). On top of the LFP trace, a simultaneously occurring Up state (as determined using multi-unit spike data from all orbitofrontal tetrodes in that session) is symbolized by the rectangular excursion. At the right-hand bottom side, spindle activity recorded from OFC is plotted in register with a ripple simultaneously recorded from hippocampal area CA1 (total duration: 0.32 s), in line with the proposal that thalamocortical spindle activity organizes a nested “hippocampal loop” (mediated by sharp-wave ripples) in time. Hippocampal ripples provide temporal structure to local spike sequences in cell assemblies, represented by colored dots plotted alongside a ripple in an hourglass (each color corresponds to a cell; data from Lansink et al., 2009). These temporally concentrated spike patterns reach target areas of the hippocampus showing reactivation as well, such as the sensory cortices, amygdala, ventral striatum, and VTA. Some of the projections in this network have been omitted for clarity (e.g., PFC-basal ganglia projections; sensory to hippocampal projections). Recordings are taken from unpublished data (Rusu, Joëls, and Pennartz)

temporal “organizer” or sequencer of representations (cf. Eichenbaum, 2013), which does not contradict its function in feeding world-state information into its target areas. Sequences internally generated in the HPC (Pastalkova et al., 2008; Pezzulo et al., 2014) may temporally align and bind single-item memory representations in cortical and sub-cortical networks together. We conclude this section with a brief reflection on the arguments supporting a hierarchical organization for sequential GDB, as opposed to nonhierarchical structures such as an undifferentiated parallel-distributed network (Cleeremans et al., 1998; Elman, 1990). Apart from the behavioral and computational arguments raised in the Introduction, we have argued that the progression from prefrontal–ventral basal ganglia loops (involved in long-term, homeostatic goals) toward the caudal sensorimotor cortico-basal ganglia loops (involved in short-term actions and habits) basically mirrors the hierarchical structure of complex sequential behavior. This mapping from behavior onto neural substrates (Figure 2) is supported by the strong projections from the HPC, amygdala, and ventromedial dopaminergic system to prefrontal-ventral basal ganglia loops, providing them with world-state, prospective and value-related information subserving GDB, whereas the DLS and its connected basal ganglia structures are largely devoid of these inputs and are supplied instead with inputs from the sensorimotor cortices and lateral dopaminergic cell groups, associated with habit formation, sensorimotor saliency, and surprise. Further reinforcing this concept of hierarchical

organization, the ventromedial DA cells exert control over more dorsal and lateral regions of the striatum, whereas no projections are known in the opposite direction, that is, from lateral dopaminergic cells to the VS. Finally, the proposed frontal-to-caudal hierarchy is markedly paralleled by the traveling direction of slow waves during non-REM sleep, which appear to play a role in the overall organization and initiation of memory consolidation. More evidence is needed to substantiate this proposal and we emphasize the likelihood of additional bidirectional information trafficking between frontal and caudal loops.

## 8 | CONCLUDING REMARKS AND PREDICTIONS FROM THEORY

Our synthesis of observations, sometimes derived from findings that may seem unconnected at first glance, can be summarized as follows. The main brain structures reviewed here (PFC, HPC, striatum, dopaminergic mesencephalon) show accumulating evidence for roles in MBL and/or GDB, taking note of the exception that the DLS is more firmly associated with habit formation. Although evidence for MBL and GDB is often jointly found for the same structure (such as mPFC, OFC, and DMS), these functions do not always or necessarily coincide (e.g., the HPC shows evidence for MBL but not for GDB; VS shows evidence for MBL while it is more generally involved in motivational control than exclusively in GDB or HB). Furthermore, the functions of these

structures are more complex than is captured by the two dichotomies under scrutiny. For instance, whereas the HPC is proposed to represent world states, store this information in support of episodic memory, and facilitate planned behavior by (re)generating state sequences, the PFC is implied in the coding of a task space in which stimuli, actions, and outcomes are interconnected via task rules. The midbrain dopaminergic system has been involved in MBL in addition to MFL, and its facilitation of GDB versus HB probably depends on the brain structure targeted by a specific dopaminergic projection. In addition to signaling RPEs, this system more generally subserves behavioral reactivity to surprising sensory inputs, movement initiation, and response invigoration. When considering these brain structures altogether, three main components of a system for hierarchical behavioral control can be distinguished: (a) a high-level, affective network for MBL and behavior directed toward long-term goals, consisting of the omPFC-VS loop with inputs from HPC, amygdala, and ventromedial DA cells; (b) a medium-level, more cognitively oriented network mediating short-term tasks in service of reaching long-term goals, consisting of the dmPFC-DMS loop, and (c) a low-level motor network implementing habitual subroutines, consisting of loops comprising the sensorimotor cortices projecting to the DLS. In these interconnected loops, dopaminergic neurons are vigorously activated by unexpected inputs, facilitating switches to novel GDB in case of changes in reward prediction, or adjustments of low-level subroutines in case of sensorimotor surprise. Evidence for memory reprocessing during sleep and other offline periods suggests a frontal-to-caudal directionality in cortico-thalamic-hippocampal circuits supporting memory consolidation, in line with the directionality proposed for on-line behavioral control.

Despite the evidence raised in support of the hypotheses proposed above, many more studies will be needed to test them in detail. The hypotheses make specific predictions to guide such tests; here we are limited to mentioning only a few. First, more electrophysiological studies are needed to determine whether coding of outcome relates to GDB and/or MBL, using outcome devaluation or switches between rewards with equal value but different sensory quality. We predict that subsets of PFC, VS, DMS, and HPC neurons will show sensitivity to outcome devaluation (conforming to GDB), but also to sensory-specific outcome properties (conforming to MBL). These types of sensitivity should be found less frequently in the DLS and other components of cache-based motor loops. For DA neurons, it will be interesting to shift away from relatively passive tasks such as Pavlovian conditioning, and re-examine their role in complex motor behaviors, particularly spontaneous, on-the-fly behaviors and behaviors requiring deliberation and prospection. Here we predict that DA neurons will causally contribute to complex behaviors involving GDB and MBL.

A second observation is that the computational concept of MBL is not congruent with the behavioral notion of GDB. The point in case is that the HPC is not required to express action–outcome relationships behaviorally but can nonetheless contribute to MBL by its episodic memory capacities and mapping of world-state variables. There is a strong need here to test the causal roles of HPC, PFC, and striatal

areas in GDB vis-à-vis MBL more systematically. For instance, the frontal-to-caudal directionality proposed for both hierarchical behavioral control and memory consolidation awaits further testing by optogenetic or chemogenetic manipulation. We predict that cortico-basal ganglia loops will causally interact with each other and that the omPFC loop exerts controls over the dmPFC loop, which in turn regulates the sensorimotor-DLS loop (with less control in the opposite direction). Specifically, inactivation of a high-level control loop will disrupt the functioning of lower-level loops, but not or less so vice versa.

Within the frontal-to-caudal organization for hierarchical control of both overt behaviors and memory consolidation, the system's front end is predicted to represent and store task-space information in its synaptic matrices, which can be investigated by blocking synaptic plasticity mechanisms during task acquisition and adaptation to altered contingencies (cf. van Wingerden et al., 2012). The HPC is expected to feed this prefrontal system with world-state information in a communication mode characterized by theta phase precession that organizes mnemonic sequencing in time. Vice versa, also prospective activity in HPC is predicted to depend on other structures, in particular on prefrontal–thalamic activity (Schmidt et al., 2019). Thus, we predict a causal role of PFC–thalamic circuits in structuring the timing of hippocampal SWRs and replay, which in turn coordinate replay in hippocampal target structures. Thus, disruption of PFC–thalamic spindle activity is predicted to disorganize ripple-associated replay in HPC and its target areas.

## ACKNOWLEDGMENTS

We wish to thank Yasmin Mzayek for commenting on the manuscript and for helping with the illustrations. We thank Hanna Bodde for editing support. This publication was supported by the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement No. 720720 (Human Brain Project SGA1) and No. 785907 (Human Brain Project SGA2), and by the Netherlands Organization for Scientific Research (NWO) under Grant No. 823.02.020 (ALW Open Competition).

## CONFLICTS OF INTEREST

None declared.

## Data availability statement

Data sharing not applicable – no new data generated.

## ORCID

Cyriel M. A. Pennartz  <https://orcid.org/0000-0001-8328-1175>

## REFERENCES

- Agster, K. L., Fortin, N. J., & Eichenbaum, H. (2002). The hippocampus and disambiguation of overlapping sequences. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 22, 5760–5768.
- Alexander, G. E., Crutcher, M. D., & DeLong, M. R. (1990). Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Progress in Brain Research*, 85, 119–146.
- Allen, T. A., Salz, D. M., McKenzie, S., & Fortin, N. J. (2016). Nonspatial sequence coding in CA1 neurons. *The Journal of Neuroscience*, 36, 1547–1563.
- Allen, W. E., Kauvar, I. V., Chen, M. Z., Richman, E. B., Yang, S. J., Chan, K., ... Deisseroth, K. (2017). Global representations of goal-directed behavior in distinct cell types of mouse neocortex. *Neuron*, 94, 891–907.e6.
- Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature*, 543, 719.
- Azuar, C., Reyes, P., Slachevsky, A., Volle, E., Kinkingnehun, S., Kouneiher, F., ... Levy, R. (2014). Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions. *NeuroImage*, 84, 1053–1060.
- Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in Cognitive Sciences*, 22, 170–188.
- Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., & Jung, M. W. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron*, 40, 177–188.
- Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37, 407–419.
- Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, 35, 48–69.
- Bao, S., Chan, V. T., & Merzenich, M. M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature*, 412, 79–83.
- Barbey, A. K., Krueger, F., & Grafman, J. (2009). Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364, 1291–1300.
- Bar-Gad, I., Morris, G., & Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in Neurobiology*, 71, 439–473.
- Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, 108, 44–79.
- Barker, G. R. I., Banks, P. J., Scott, H., Ralph, G. S., Mitrophanous, K. A., Wong, L.-F., ... Warburton, E. C. (2017). Separate elements of episodic memory subserved by distinct hippocampal-prefrontal connections. *Nature Neuroscience*, 20, 242–250.
- Barto, A. G. (1994). Reinforcement learning control. *Current Opinion in Neurobiology*, 4, 888–893.
- Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, 13, 341–379.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76, 695–711.
- Battaglia, F. P., Benchenane, K., Sirota, A., Pennartz, C. M. A., & Wiener, S. I. (2011). The hippocampus: Hub of brain network communication for memory. *Trends in Cognitive Sciences*, 15, 310–318.
- Battaglia, F. P., & Pennartz, C. M. (2011). The construction of semantic memory: Grammar-based representations learned from relational episodic information. *Frontiers in Computational Neuroscience*, 5, 36.
- Belin-Rauscent, A., Everitt, B. J., & Belin, D. (2012). Intrastratial shifts mediate the transition from drug-seeking actions to habits. *Biological Psychiatry*, 72, 343–345.
- Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., & Wiener, S. I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron*, 66, 921–936.
- Berendse, H. W., Groenewegen, H. J., & Lohman, A. H. (1992). Compartmental distribution of ventral striatal neurons projecting to the mesencephalon in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 12, 2079–2103.
- Berke, J. D. (2018). What does dopamine mean? *Nature Neuroscience*, 21, 787–793.
- Berke, J. D., Okatan, M., Skurski, J., & Eichenbaum, H. B. (2004). Oscillatory entrainment of striatal neurons in freely moving rats. *Neuron*, 43, 883–896.
- Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology*, 191, 391–431.
- Bett, D., Allison, E., Murdoch, L. H., Kaefer, K., Wood, E. R., & Dudchenko, P. A. (2012). The neural substrates of deliberative decision making: Contrasting effects of hippocampus lesions on performance and vicarious trial-and-error behavior in a spatial memory task and a visual discrimination task. *Frontiers in Behavioral Neuroscience*, 6, 70.
- Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 20, 4320–4324.
- Bos, J. J., Vinck, M., van Mourik-Donga, L. A., Jackson, J. C., Witter, M. P., & Pennartz, C. M. A. (2017). Perirhinal firing patterns are sustained across large spatial segments of the task environment. *Nature Communications*, 8, 15602.
- Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, 12, 201–208.
- Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113, 262–280.
- Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron*, 68, 815–834.
- Bugaysen, J., Bar-Gad, I., & Korngreen, A. (2013). Continuous modulation of action potential firing by a unitary GABAergic connection in the globus pallidus in vitro. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33, 12805–12809.
- Burke, D. A., Rotstein, H. G., & Alvarez, V. A. (2017). Striatal local circuitry: A new framework for lateral inhibition. *Neuron*, 96, 267–284.
- Burke, K. A., Takahashi, Y. K., Correll, J., Brown, P. L., & Schoenbaum, G. (2009). Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with 'disinhibition' of responding for previously unrewarded cues. *The European Journal of Neuroscience*, 30, 1941–1946.
- Buzsáki, G. (1986). Hippocampal sharp waves: Their origin and significance. *Brain Research*, 398, 242–252.
- Buzsáki, G. (1989). Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience*, 31, 551–570.
- Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16, 130–138.
- Cabral, H. O., Vinck, M., Fouquet, C., Pennartz, C. M., Rondi-Reig, L., & Battaglia, F. P. (2014). Oscillatory dynamics and place field maps reflect hippocampal ensemble processing of sequence and place memory under NMDA receptor control. *Neuron*, 81, 402–415.
- Calabresi, P., Picconi, B., Tozzi, A., & Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neurosciences*, 30, 211–219.
- Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 26, 321–352.

- Cardinal, R. N., Pennicott, D. R., Sugathapala, C. L., Robbins, T. W., & Everitt, B. J. (2001). Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science (New York, N.Y.)*, *292*, 2499–2501.
- Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *The Journal of Comparative Neurology*, *371*, 179–207.
- Chiang, F. K., & Wallis, J. D. (2018). Neuronal encoding in prefrontal cortex during hierarchical reinforcement learning. *Journal of Cognitive Neuroscience*, *30*, 1197–1208.
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406–416.
- Cooch, N. K., Stalnaker, T. A., Wied, H. M., Bali-Chaudhary, S., McDannald, M. A., Liu, T. L., & Schoenbaum, G. (2015). Orbitofrontal lesions eliminate signalling of biological significance in cue-responsive ventral striatal neurons. *Nature Communications*, *6*, 7195.
- Corbit, L. H., & Balleine, B. W. (2000). The role of the hippocampus in instrumental conditioning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*, 4233–4239.
- Corbit, L. H., & Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research*, *146*, 145–157.
- Corbit, L. H., Muir, J. L., & Balleine, B. W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *The European Journal of Neuroscience*, *18*, 1286–1294.
- Corbit, L. H., Ostlund, S. B., & Balleine, B. W. (2002). Sensitivity to instrumental contingency degradation is mediated by the entorhinal cortex and its efferents via the dorsal hippocampus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*, 10976–10984.
- Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews Neuroscience*, *3*, 153–160.
- Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research*, *146*, 167–174.
- Cowen, S. L., Davis, G. A., & Nitz, D. A. (2012). Anterior cingulate neurons in the rat map anticipated effort and reward to their associated action sequences. *Journal of Neurophysiology*, *107*, 2393–2407.
- Critchley, H. D., & Rolls, E. T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *Journal of Neurophysiology*, *75*, 1673–1686.
- Crunelli, V., & Hughes, S. W. (2010). The slow (<1 Hz) rhythm of non-REM sleep: A dialogue between three cardinal oscillators. *Nature Neuroscience*, *13*, 9–17.
- Dalley, J. W., Cardinal, R. N., & Robbins, T. W. (2004). Prefrontal executive and cognitive functions in rodents: Neural and neurochemical substrates. *Neuroscience and Biobehavioral Reviews*, *28*, 771–784.
- Darvas, M., Wunsch, A. M., Gibbs, J. T., & Palmiter, R. D. (2014). Dopamine dependency for acquisition and performance of Pavlovian conditioned response. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 2764–2769.
- Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, *63*, 497–507.
- Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130478.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1704–1711.
- de Borchgrave, R., Rawlins, J. N., Dickinson, A., & Balleine, B. W. (2002). Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. *Experimental Brain Research*, *144*, 50–68.
- De Bruin, J. P. C., Feenstra, M. G. P., Broersen, L. M., Van Leeuwen, M., Arens, C., De Vries, S., & Joosten, R. N. (2000). Role of the prefrontal cortex of the rat in learning and decision making: Effects of transient inactivation. *Progress in Brain Research*, *126*, 103–113.
- DeLong, M. R., Crutcher, M. D., & Georgopoulos, A. P. (1983). Relations between movement and single cell discharge in the substantia nigra of the behaving monkey. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *3*, 1599–1606.
- Dembrow, N. C., Zemelman, B. V., & Johnston, D. (2015). Temporal dynamics of L5 dendrites in medial prefrontal cortex regulate integration versus coincidence detection of afferent inputs. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*, 4501–4514.
- Denenberg, V. H., Kim, D. S., & Palmiter, R. D. (2004). The role of dopamine in learning, memory, and performance of a water escape task. *Behavioural Brain Research*, *148*, 73–78.
- Desrochers, T. M., Chatham, C. H., & Badre, D. (2015). The necessity of rostralateral prefrontal cortex for higher-level sequential behavior. *Neuron*, *87*, 1357–1368.
- Dezfouli, A., Lingawi, N. W., & Balleine, B. W. (2014). Habits as action sequences: Hierarchical action control and changes in outcome value. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130482.
- Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 2733–2742.
- Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, *12*, 961–974.
- Durstewitz, D., Vitoz, N. M., Floresco, S. B., & Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, *66*, 438–448.
- Ego-Stengel, V., & Wilson, M. A. (2010). Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, *20*, 1–10.
- Eichenbaum, H. (2013). Memory on time. *Trends in Cognitive Sciences*, *17*, 81–88.
- Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience*, *15*, 732–744.
- Eichenbaum, H. (2017). Prefrontal-hippocampal interactions in episodic memory. *Nature Reviews Neuroscience*, *18*, 547–558.
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, *23*, 209–226.
- Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neuroscience and Biobehavioral Reviews*, *36*, 1597–1608.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
- Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., & Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. *Nature*, *525*, 243–246.
- Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*, 1057–1070.
- Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science (New York, N.Y.)*, *318*, 1147–1150.
- Faure, A., Haberland, U., Conde, F., & El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*, 2771–2780.
- Fisher, S. D., Robertson, P. B., Black, M. J., Redgrave, P., Sagar, M. A., Abraham, W. C., & Reynolds, J. N. J. (2017). Reinforcement determines the timing dependence of corticostriatal synaptic plasticity in vivo. *Nature Communications*, *8*, 334.

- Flaherty, A. W., & Graybiel, A. M. (1995). Motor and somatosensory corticostriatal projection magnifications in the squirrel monkey. *Journal of Neurophysiology*, *74*, 2638–2648.
- Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*, 458–462.
- Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*, 680–683.
- Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*, 119–130.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.
- Fujisawa, S., & Buzsáki, G. (2011). A 4-Hz oscillation adaptively synchronizes prefrontal, VTA and hippocampal activities. *Neuron*, *72*, 153–165.
- Fuster, J. M. (2001). The prefrontal cortex—an update: Time is of the essence. *Neuron*, *30*, 319–333.
- Gabbott, P., Warner, T. A., Brown, J., Salway, P., Gabbott, T., & Busby, S. (2012). Amygdala afferents monosynaptically innervate corticospinal neurons in rat medial prefrontal cortex. *The Journal of Comparative Neurology*, *520*, 2440–2458.
- Geisler, S., Derst, C., Veh, R. W., & Zahm, D. S. (2007). Glutamatergic afferents of the ventral tegmental area in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*, 5730–5743.
- Genovesio, A., Tsujimoto, S., & Wise, S. P. (2012). Encoding goals but not abstract magnitude in the primate prefrontal cortex. *Neuron*, *74*, 656–662.
- Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*, 1222–1223.
- Girardeau, G., Inema, I., & Buzsáki, G. (2017). Reactivations of emotional memory in the hippocampus-amygdala system during sleep. *Nature Neuroscience*, *20*, 1634–1642.
- Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., ... Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 14681–14686.
- Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*, 585–595.
- Gmaz, J. M., Carmichael, J. E., & van der Meer, M. A. (2018). Persistent coding of outcome-predictive cue features in the rat nucleus accumbens. *eLife*, *7*, e37275.
- Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *351*, 1445–1453.
- Goltstein, P. M., Coffey, E. B., Roelfsema, P. R., & Pennartz, C. M. (2013). In vivo two-photon Ca<sup>2+</sup> imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. *The Journal of Neuroscience*, *33*, 11540–11555.
- Goltstein, P. M., Meijer, G. T., & Pennartz, C. M. (2018). Conditioning sharpens the spatial representation of rewarded stimuli in mouse primary visual cortex. *eLife*, *7*, e37683.
- Gomperts, S. N., Kloosterman, F., & Wilson, M. A. (2015). VTA neurons coordinate with the hippocampal reactivation of spatial experience. *eLife*, *4*, e05360.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*, 181–197.
- Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*, 2264.
- Groenewegen, H. J., Berendse, H. W., & Haber, S. N. (1993). Organization of the output of the ventral striatopallidal system in the rat: Ventral pallidal efferents. *Neuroscience*, *57*, 113–142.
- Groenewegen, H. J., Berendse, H. W., Wolters, J. G., & Lohman, A. H. (1990). The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: Evidence for a parallel organization. *Progress in Brain Research*, *85*, 95–116 discussion 116–118.
- Groenewegen, H. J., & Uylings, H. B. (2000). The prefrontal cortex and the integration of sensory, limbic and autonomic information. *Progress in Brain Research*, *126*, 3–28.
- Groenewegen, H. J., Vermeulen-Van der Zee, E., te Kortschot, A., & Witter, M. P. (1987). Organization of the projections from the subiculum to the ventral striatum in the rat. A study using anterograde transport of *Phaseolus vulgaris* Leucoagglutinin. *Neuroscience*, *23*, 103–120.
- Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*, 2369–2382.
- Haddon, J. E., & Killcross, S. (2006). Prefrontal cortex lesions disrupt the contextual control of response conflict. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*, 2933–2940.
- Hagan, J. J., Alpert, J. E., Morris, R. G., & Iversen, S. D. (1983). The effects of central catecholamine depletions on spatial learning in rats. *Behavioural Brain Research*, *9*, 83–104.
- Hall, J., Parkinson, J. A., Connor, T. M., Dickinson, A., & Everitt, B. J. (2001). Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. *The European Journal of Neuroscience*, *13*, 1984–1992.
- Handelmann, G. E., & Olton, D. S. (1981). Spatial memory following damage to hippocampal CA3 pyramidal cells with kainic acid: Impairment and recovery with preoperative training. *Brain Research*, *217*, 41–58.
- Hansen, N., & Manahan-Vaughan, D. (2014). Dopamine D1/D5 receptors mediate informational saliency that promotes persistent hippocampal long-term plasticity. *Cerebral Cortex (New York, NY: 1991)*, *24*, 845–858.
- Hart, G., Bradfield, L. A., & Balleine, B. W. (2018). Prefrontal cortico-striatal disconnection blocks the acquisition of goal-directed action. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(5), 1311–1322.
- Hart, G., Leung, B. K., & Balleine, B. W. (2014). Dorsal and ventral streams: The distinct role of striatal subregions in the acquisition and performance of goal-directed actions. *Neurobiology of Learning and Memory*, *108*, 104–118.
- Harvey, C. D., Coen, P., & Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, *484*, 62–68.
- Heilbronner, S. R., Rodriguez-Romaguera, J., Quirk, G. J., Groenewegen, H. J., & Haber, S. N. (2016). Circuit-based Corticostriatal homologies between rat and primate. *Biological Psychiatry*, *80*, 509–521.
- Hitchcott, P. K., Quinn, J. J., & Taylor, J. R. (2007). Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. *Cerebral Cortex (New York, NY: 1991)*, *17*, 2820–2827.
- Hok, V., Save, E., Lenck-Santini, P. P., & Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 4602–4607.

- Holland, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, 30, 104–117.
- Hoover, W. B., & Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure & Function*, 212, 149–179.
- Hosokawa, T., Kennerley, S. W., Sloan, J., & Wallis, J. D. (2013). Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33, 17385–17397.
- 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* (pp. 249–270). Cambridge, MA: MIT Press.
- Howard, C. D., Li, H., Geddes, C. E., & Jin, X. (2017). Dynamic nigrostriatal dopamine biases action selection. *Neuron*, 93, 1436–1450.e8.
- Huguenard, J. R., & McCormick, D. A. (2007). Thalamic synchrony and dynamic regulation of global forebrain oscillations. *Trends in Neurosciences*, 30, 350–356.
- Huxter, J., Burgess, N., & O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. *Nature*, 425, 828–832.
- Inglis, W. L., & Winn, P. (1995). The pedunculopontine tegmental nucleus: Where the striatum meets the reticular formation. *Progress in Neurobiology*, 47, 1–29.
- Ito, H. T., Zhang, S. J., Witter, M. P., Moser, E. I., & Moser, M. B. (2015). A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature*, 522, 50–55.
- Ito, M., & Doya, K. (2015). Distinct neural representation in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35, 3499–3514.
- Ito, R., Robbins, T. W., Pennartz, C. M., & Everitt, B. J. (2008). Functional interaction between the hippocampus and nucleus accumbens shell is necessary for the acquisition of appetitive spatial context conditioning. *The Journal of Neuroscience*, 28, 6950–6959.
- Jackson, S. A., Horst, N. K., Pears, A., Robbins, T. W., & Roberts, A. C. (2016). Role of the perigenual anterior cingulate and orbitofrontal cortex in contingency learning in the marmoset. *Cerebral Cortex (New York, NY: 1991)*, 26, 3273–3284.
- Jadhav, S. P., Kemere, C., German, P. W., & Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science (New York, N.Y.)*, 336, 1454–1458.
- Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, 10, 100–107.
- Jimura, K., Chushak, M. S., & Braver, T. S. (2013). Impulsivity and self-control during intertemporal decision making linked to the neural dynamics of reward value representation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33, 344–357.
- Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience*, 27, 12176–12189.
- Johnson, L. A., Blakely, T., Hermes, D., Hakimian, S., Ramsey, N. F., & Ojemann, J. G. (2012). Sleep spindles are locally modulated by training on a brain-computer interface. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 18583–18588.
- Jones, J. L., Esber, G. R., McDannald, M. A., Gruber, A. J., Hernandez, A., Mirenzi, A., & Schoenbaum, G. (2012). Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science (New York, N.Y.)*, 338, 953–956.
- Jones, M. W., & Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biology*, 3, e402.
- Kalenscher, T., & Pennartz, C. M. (2008). Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. *Progress in Neurobiology*, 84, 284–315.
- Kargo, W. J., Szatmary, B., & Nitz, D. A. (2007). Adaptation of prefrontal cortical firing patterns and their fidelity to changes in action-reward contingencies. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27, 3548–3559.
- Keller, G. B., Bonhoeffer, T., & Hubener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, 74, 809–815.
- Kelley, A. E. (2004). Ventral striatal control of appetitive motivation: Role in ingestive behavior and reward-related learning. *Neuroscience and Biobehavioral Reviews*, 27, 765–776.
- Kesner, R. P., Hunsaker, M. R., & Warthen, M. W. (2008). The CA3 subregion of the hippocampus is critical for episodic memory processing by means of relational encoding in rats. *Behavioral Neuroscience*, 122, 1217–1225.
- Kim, H. F., & Hikosaka, O. (2013). Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron*, 79, 1001–1010.
- Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., ... Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science (New York, N.Y.)*, 356, 73–78.
- Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science (New York, N.Y.)*, 318, 594–598.
- Kraus, B. J., Robinson, R. J., 2nd, White, J. A., Eichenbaum, H., & Hasselmo, M. E. (2013). Hippocampal "time cells": Time versus path integration. *Neuron*, 78, 1090–1101.
- Kudrimoti, H. S., Barnes, C. A., & McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: Effects of behavioral state, experience, and EEG dynamics. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 19, 4090–4101.
- Kyd, R. J., & Bilkey, D. K. (2003). Prefrontal cortex lesions modify the spatial properties of hippocampal place cells. *Cerebral Cortex*, 13, 444–451.
- Ladenbauer, J., Ladenbauer, J., Kulzow, N., de Boer, R., Avramova, E., Grittner, U., & Floel, A. (2017). Promoting sleep oscillations and their functional coupling by transcranial stimulation enhances memory consolidation in mild cognitive impairment. *The Journal of Neuroscience*, 37, 7111–7124.
- Lalla, L., Rueda Orozco, P. E., Jurado-Parras, M. T., Brovelli, A., & Robbe, D. (2017). Local or not local: Investigating the nature of striatal Theta oscillations in behaving rats. *eNeuro*, 4, e0128–17.2017.
- Lansink, C. S., Goltstein, P. M., Lankelma, J. V., Joosten, R. N., McNaughton, B. L., & Pennartz, C. M. A. (2008). Preferential reactivation of motivationally relevant information in the ventral striatum. *The Journal of Neuroscience*, 28, 6372–6382.
- Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., & Pennartz, C. M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biology*, 7, e1000173.
- Lansink, C. S., Goltstein, P. M., Lankelma, J. V., & Pennartz, C. M. (2010). Fast-spiking interneurons of the rat ventral striatum: Temporal coordination of activity with principal cells and responsiveness to reward. *The European Journal of Neuroscience*, 32, 494–508.
- Lansink, C. S., Jackson, J. C., Lankelma, J. V., Ito, R., Robbins, T. W., Everitt, B. J., & Pennartz, C. M. (2012). Reward cues in space: Commonalities and differences in neural coding by hippocampal and ventral striatal ensembles. *The Journal of Neuroscience*, 32, 12444–12459.
- Lansink, C. S., Meijer, G. T., Lankelma, J. V., Vinck, M. A., Jackson, J. C., & Pennartz, C. M. A. (2016). Reward expectancy strengthens CA1 Theta and Beta band synchronization and hippocampal-ventral striatal coupling. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36, 10598–10610.

- Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Hixon symposium on cerebral mechanisms in behavior*. New York: Wiley.
- Latchoumane, C.-F. V., Ngo, H.-V. V., Born, J., & Shin, H.-S. (2017). Thalamic spindles promote memory formation during sleep through triple phase-locking of cortical, thalamic, and hippocampal rhythms. *Neuron*, 95, 424–435.e6.
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 20, 1434–1448.
- Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A., & Keller, G. B. (2017). A sensorimotor circuit in mouse cortex for visual flow predictions. *Neuron*, 95(1420–1432), e1425.
- Luthi, A. (2014). Sleep spindles: Where they come from, what they do. *The Neuroscientist*, 20, 243–256.
- MacDonald, C. J., Carrow, S., Place, R., & Eichenbaum, H. (2013). Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33, 14607–14616.
- MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal “time cells” bridge the gap in memory for discontinuous events. *Neuron*, 71, 737–749.
- Maingret, N., Girardeau, G., Todorova, R., Goutier, M., & Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. *Nature Neuroscience*, 19, 959.
- Mar, A. C., Walker, A. L., Theobald, D. E., Eagle, D. M., & Robbins, T. W. (2011). Dissociable effects of lesions to orbitofrontal cortex subregions on impulsive choice in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31, 6398–6404.
- Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, 15, 238–300.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 262, 23–81.
- Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature*, 444, 610–613.
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S., & Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *The Journal of Neuroscience*, 24, 6862–6870.
- Maurin, Y., Banrezes, B., Menetrey, A., Maily, P., & Deniau, J. M. (1999). Three-dimensional distribution of nigrostriatal neurons in the rat: Relation to the topography of striatonigral projections. *Neuroscience*, 91, 891–909.
- McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31, 2700–2705.
- McDannald, M. A., Takahashi, Y. K., Lopatina, N., Pietras, B. W., Jones, J. L., & Schoenbaum, G. (2012). Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. *The European Journal of Neuroscience*, 35, 991–996.
- McNamara, C. G., Tejero-Cantero, Á., Trouche, S., Campo-Urriza, N., & Dupret, D. (2014). Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature Neuroscience*, 17, 1658.
- Meck, W. H. (1988). Hippocampal function is required for feedback control of an internal clock's criterion. *Behavioral Neuroscience*, 102, 54–60.
- Meck, W. H., Church, R. M., & Olton, D. S. (2013). Hippocampus, time, and memory. *Behavioral Neuroscience*, 127, 655–668.
- Menegas, W., Bergan, J. F., Ogawa, S. K., Isogai, Y., Umadevi Venkataraju, K., Osten, P., ... Watabe-Uchida, M. (2015). Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. *eLife*, 4, e10032.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York, NY: Henry Holt and Co.
- Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, 20, 445–468.
- Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, 72, 1024–1027.
- Miyamoto, D., Hirai, D., Fung, C. C., Inutsuka, A., Odagawa, M., Suzuki, T., et al. (2016). Top-down cortical input during NREM sleep consolidates perceptual memory. *Science (New York, N.Y.)*, 352, 1315–1318.
- Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology*, 14, 69–97.
- Molle, M., & Born, J. (2011). Slow oscillations orchestrating fast oscillations and memory consolidation. *Progress in Brain Research*, 193, 93–110.
- Morris, R. G., Anderson, E., Lynch, G. S., & Baudry, M. (1986). Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature*, 319, 774–776.
- Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., et al. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, 207, 35–66.
- Mulder, A. B., Nordquist, R. E., Orgut, O., & Pennartz, C. M. (2003). Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. *Behavioural Brain Research*, 146, 77–88.
- Naneix, F., Marchand, A. R., Di Scala, G., Pape, J. R., & Coutureau, E. (2009). A role for medial prefrontal dopaminergic innervation in instrumental conditioning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29, 6599–6606.
- Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., & Tononi, G. (2011). Regional slow waves and spindles in human sleep. *Neuron*, 70(1), 153–169.
- Nonomura, S., Nishizawa, K., Sakai, Y., Kawaguchi, Y., Kato, S., Uchigashima, M., et al. (2018). Monitoring and updating of action selection for goal-directed behavior through the striatal direct and indirect pathways. *Neuron*, 99, 1302–1314.e5.
- Nordquist, R. E., Voorn, P., de Mooij-van Malsen, J. G., Joosten, R. N., Pennartz, C. M., & Vanderschuren, L. J. (2007). Augmented reinforcer value and accelerated habit formation after repeated amphetamine treatment. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, 17, 532–540.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- O'Neill, P. K., Gordon, J. A., & Sigurdsson, T. (2013). Theta oscillations in the medial prefrontal cortex are modulated by spatial working memory and synchronize with the hippocampus through its ventral subregion. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33, 14211–14224.
- O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, 18, 283–328.
- Ostlund, S. B., & Balleine, B. W. (2007). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27, 4819–4825.
- Padoa-Schioppa, C., & Conen, K. E. (2017). Orbitofrontal cortex: A neural circuit for economic decisions. *Neuron*, 96, 736–754.

- Papale, A. E., Zielinski, M. C., Frank, L. M., Jadhav, S. P., & Redish, A. D. (2016). Interplay between hippocampal sharp-wave-ripple events and vicarious trial and error behaviors in decision making. *Neuron*, *92*, 975–982.
- Parkes, S. L., Ravassard, P. M., Cerpa, J. C., Wolff, M., Ferreira, G., & Coutureau, E. (2018). Insular and ventrolateral orbitofrontal cortices differentially contribute to goal-directed behavior in rodents. *Cerebral Cortex (New York, NY: 1991)*, *28*, 2313–2325.
- Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science (New York, N.Y.)*, *321*, 1322–1327.
- Pawlak, V., & Kerr, J. N. (2008). Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*, 2435–2446.
- Pennartz, C. M. (1997). Reinforcement learning by Hebbian synapses with adaptive thresholds. *Neuroscience*, *81*, 303–319.
- Pennartz, C. M., Ameerun, R. F., Groenewegen, H. J., & Lopes da Silva, F. H. (1993). Synaptic plasticity in an in vitro slice preparation of the rat nucleus accumbens. *The European Journal of Neuroscience*, *5*, 107–117.
- Pennartz, C. M., Berke, J. D., Graybiel, A. M., Ito, R., Lansink, C. S., van der Meer, M., ... Voorn, P. (2009). Corticostriatal interactions during learning, memory processing, and decision making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*, 12831–12838.
- Pennartz, C. M., Groenewegen, H. J., & Lopes da Silva, F. H. (1994). The nucleus accumbens as a complex of functionally distinct neuronal ensembles: An integration of behavioural, electrophysiological and anatomical data. *Progress in Neurobiology*, *42*, 719–761.
- Pennartz, C. M., Lee, E., Verheul, J., Lipa, P., Barnes, C. A., & McNaughton, B. L. (2004). The ventral striatum in off-line processing: Ensemble reactivation during sleep and modulation by hippocampal ripples. *The Journal of Neuroscience*, *24*, 6446–6456.
- Pennartz, C. M., van Wingerden, M., & Vinck, M. (2011). Population coding and neural rhythmicity in the orbitofrontal cortex. *Annals of the New York Academy of Sciences*, *1239*, 149–161.
- Pennartz, C. M. A. (1996). The ascending neuromodulatory systems in learning by reinforcement: Comparing computational conjectures with experimental findings. *Brain Research. Brain Research Reviews*, *21*, 219–245.
- Pennartz, C. M. A. (2015). *The brain's representational power—On consciousness and the integration of modalities*. Cambridge, MA: MIT Press.
- Pennartz, C. M. A. (2018). Consciousness, representation, action: The importance of being goal-directed. *Trends in Cognitive Sciences*, *22*, 137–153.
- Pennartz, C. M. A., Ito, R., Verschure, P. F., Battaglia, F. P., & Robbins, T. W. (2011). The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends in Neurosciences*, *34*, 548–559.
- Penny, W. D., Zeidman, P., & Burgess, N. (2013). Forward and backward inference in spatial cognition. *PLoS Computational Biology*, *9*, e1003383.
- Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, *12*(7), 919.
- Pezzulo, G., van der Meer, M. A. A., Lansink, C. S., & Pennartz, C. M. A. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, *18*, 647–657.
- Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*, 74–79.
- Plenz, D. (2003). When inhibition goes incognito: Feedback interaction between spiny projection neurons in striatal function. *Trends in Neurosciences*, *26*, 436–443.
- Preuss, T. M. (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. *Journal of Cognitive Neuroscience*, *7*, 1–24.
- Qin, Y. L., McNaughton, B. L., Skaggs, W. E., & Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*, 1525–1533.
- Ramus, S. J., & Eichenbaum, H. (2000). Neural correlates of olfactory recognition memory in the rat orbitofrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*, 8199–8208.
- Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*, *20*, 1263–1290.
- Redgrave, P., Prescott, T. J., & Gurney, K. (1999a). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience*, *89*, 1009–1023.
- Redgrave, P., Prescott, T. J., & Gurney, K. (1999b). Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences*, *22*, 146–151.
- Redish, A. D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, *17*, 147–159.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research* (pp. 64–99). New York: Appleton-Century-Crofts.
- Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature*, *413*, 67–70.
- Rich, E. L., & Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*, 7208–7219.
- Robbins, T. W., Cador, M., Taylor, J. R., & Everitt, B. J. (1989). Limbic-striatal interactions in reward-related processes. *Neuroscience and Biobehavioral Reviews*, *13*, 155–162.
- Robbins, T. W., & Everitt, B. J. (1982). Functional studies of the central catecholamines. *International Review of Neurobiology*, *23*, 303–365.
- Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology*, *6*, 228–236.
- Rodriguez-Oroz, M. C., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E., & Obeso, J. A. (2009). Initial clinical manifestations of Parkinson's disease: Features and pathophysiological mechanisms. *Lancet Neurology*, *8*, 1128–1139.
- Roesch, M. R., Singh, T., Brown, P. L., Mullins, S. E., & Schoenbaum, G. (2009). Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*, 13365–13376.
- Roitman, M. F., Wheeler, R. A., & Carelli, R. M. (2005). Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron*, *45*, 587–597.
- Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex (New York, NY: 1991)*, *10*, 284–294.
- Rothschild, G., Eban, E., & Frank, L. M. (2017). A cortical-hippocampal-cortical loop of information processing during memory consolidation. *Nature Neuroscience*, *20*, 251–259.
- Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. S. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience*, *9*, 1161–1168.
- Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, *70*, 1054–1069.
- Rusu, S.I., Bos, J.J., Lankelma, J.V., Gentet, L.J., Joëls, M., and Pennartz, C. M.A. (2016). *Neuronal ensemble reactivation in the orbitofrontal cortex*



- during sleep. In: Society for Neuroscience Annual Meeting, Society for Neuroscience, Washington, DC, p. 178.103.
- Sadek, A. R., Magill, P. J., & Bolam, J. P. (2007). A single-cell analysis of intrinsic connectivity in the rat globus pallidus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27, 6352–6362.
- Salamone, J. D., Cousins, M. S., & Bucher, S. (1994). Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behavioural Brain Research*, 65, 221–229.
- Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science (New York, N.Y.)*, 310, 1337–1340.
- Sato, F., Lavallee, P., Levesque, M., & Parent, A. (2000). Single-axon tracing study of neurons of the external segment of the globus pallidus in primate. *The Journal of Comparative Neurology*, 417, 17–31.
- Schmidt, B., Duin, A. A., & Redish, A. D. (2019). Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision-making. *Journal of Neurophysiology*, 121(6), 1981–2000.
- Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., & Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience*, 10, 885–892.
- Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron*, 39, 855–867.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80, 1–27.
- Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience*, 17, 183–195.
- Schultz, W., Apicella, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 12, 4595–4610.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–1599.
- Schultz, W., Ruffieux, A., & Aebischer, P. (1983). The activity of pars compacta neurons of the monkey substantia nigra in relation to motor activation. *Experimental Brain Research*, 51, 377–387.
- Schultz, W., Stauffer, W. R., & Lak, A. (2017). The phasic dopamine signal maturing: From reward via behavioural activation to formal economic utility. *Current Opinion in Neurobiology*, 43, 139–148.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11–21.
- Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron*, 60, 1169–1181.
- Sesack, S. R., & Grace, A. A. (2010). Cortico-basal ganglia reward network: Microcircuitry. *Neuropsychopharmacology*, 35, 27–47.
- Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science (New York, N.Y.)*, 321, 848–851.
- Shipman, M. L., Trask, S., Bouton, M. E., & Green, J. T. (2018). Inactivation of prelimbic and infralimbic cortex respectively affects minimally-trained and extensively-trained goal-directed actions. *Neurobiology of Learning and Memory*, 155, 164–172.
- Shuler, M. G., & Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science*, 311, 1606–1609.
- Shushruth, S., Mazurek, M., & Shadlen, M. N. (2018). Comparison of decision-related signals in sensory and motor preparatory responses of neurons in area LIP. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 38, 6350–6365.
- Simon, D. A., & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31, 5526–5539.
- Sirota, A., Csicsvari, J., Buhl, D., & Buzsaki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 2065–2069.
- Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, 79, 361–374.
- Stalnaker, T. A., Calhoun, G. G., Ogawa, M., Roesch, M. R., & Schoenbaum, G. (2010). Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Frontiers in Integrative Neuroscience*, 4, 12.
- Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nature Neuroscience*, 18, 620–627.
- Steiner, A. P., & Redish, A. D. (2014). Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nature Neuroscience*, 17, 995–1002.
- Stoianov, I. P., Pennartz, C. M. A., Lansink, C. S., & Pezzulo, G. (2018). Model-based spatial navigation in the hippocampus-ventral striatum circuit: A computational analysis. *PLoS Computational Biology*, 14, e1006316.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.
- Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., & Schoenbaum, G. (2017). Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron*, 95 (1395–1405), e1393.
- Takahashi, Y. K., Langdon, A. J., Niv, Y., & Schoenbaum, G. (2016). Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron*, 91, 182–193.
- Takehara-Nishiuchi, K., & McNaughton, B. L. (2008). Spontaneous changes of neocortical code for associative memory during consolidation. *Science (New York, N.Y.)*, 322, 960–963.
- Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33, 15376–15381.
- Tang, W., Shin, J. D., Frank, L. M., & Jadhav, S. P. (2017). Hippocampal-prefrontal reactivation during learning is stronger in awake compared with sleep states. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 37, 11789–11805.
- Taverna, S., Canciani, B., & Pennartz, C. M. (2007). Membrane properties and synaptic connectivity of fast-spiking interneurons in rat ventral striatum. *Brain Research*, 1152, 49–56.
- Taverna, S., van Dongen, Y. C., Groenewegen, H. J., & Pennartz, C. M. (2004). Direct physiological evidence for synaptic connectivity between medium-sized spiny neurons in rat nucleus accumbens in situ. *Journal of Neurophysiology*, 91, 1111–1121.
- Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100, 147–154.
- Thomas, M. J., Malenka, R. C., & Bonci, A. (2000). Modulation of long-term depression by dopamine in the mesolimbic system. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 20, 5581–5586.
- Thorn, C. A., Atallah, H., Howe, M., & Graybiel, A. M. (2010). Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron*, 66, 781–795.
- Tsai, H. C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., de Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science (New York, N.Y.)*, 324, 1080–1084.
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. (2007). Schemas and memory consolidation. *Science (New York, N.Y.)*, 316, 76–82.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

- Uylyngs, H. B., Groenewegen, H. J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research*, *146*, 3–17.
- Valdes, J. L., McNaughton, B. L., & Fellous, J. M. (2015). Offline reactivation of experience-dependent neuronal firing patterns in the rat ventral tegmental area. *Journal of Neurophysiology*, *114*, 1183–1195.
- Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*, 4019–4026.
- van der Meer, M. A., Johnson, A., Schmitzer-Torbert, N. C., & Redish, A. D. (2010). Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron*, *67*, 25–32.
- van der Meer, M. A., & Redish, A. D. (2011a). Theta phase precession in rat ventral striatum links place and reward information. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*, 2843–2854.
- van der Meer, M. A., & Redish, A. D. (2011b). Ventral striatum: A critical look at models of learning and evaluation. *Current Opinion in Neurobiology*, *21*, 387–392.
- van der Meer, M. A. A., & Redish, A. D. (2009). Covert expectation-of-reward in rat ventral striatum at decision points. *Frontiers in Integrative Neuroscience*, *3*, 1.
- van Dongen, Y. C., Deniau, J. M., Pennartz, C. M., Galis-de Graaf, Y., Voorn, P., Thierry, A. M., & Groenewegen, H. J. (2005). Anatomical evidence for direct connections between the shell and core subregions of the rat nucleus accumbens. *Neuroscience*, *136*, 1049–1071.
- van Duuren, E., Lankelma, J., & Pennartz, C. M. (2008). Population coding of reward magnitude in the orbitofrontal cortex of the rat. *The Journal of Neuroscience*, *28*, 8590–8603.
- van Duuren, E., van der Plasse, G., Lankelma, J., Joosten, R. N., Feenstra, M. G., & Pennartz, C. M. (2009). Single-cell and population coding of expected reward probability in the orbitofrontal cortex of the rat. *The Journal of Neuroscience*, *29*, 8965–8976.
- van Wingerden, M., Vinck, M., Lankelma, J., & Pennartz, C. M. (2010a). Theta-band phase locking of orbitofrontal neurons during reward expectancy. *The Journal of Neuroscience*, *30*, 7078–7087.
- van Wingerden, M., Vinck, M., Lankelma, J. V., & Pennartz, C. M. (2010b). Learning-associated gamma-band phase-locking of action-outcome selective neurons in orbitofrontal cortex. *The Journal of Neuroscience*, *30*, 10025–10038.
- van Wingerden, M., Vinck, M., Tijms, V., Ferreira, I. R., Jonker, A. J., & Pennartz, C. M. (2012). NMDA receptors control cue-outcome selectivity and plasticity of orbitofrontal firing patterns during associative stimulus-reward learning. *Neuron*, *76*, 813–825.
- Verschure, P. F., Pennartz, C. M., & Pezzulo, G. (2014). The why, what, where, when and how of goal-directed choice: Neuronal and computational principles. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1655), 20130483.
- Vickery, T. J., Chun, M. M., & Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron*, *72*, 166–177.
- Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, *27*, 468–474.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*, 953–956.
- Walton, M. E., Croxson, P. L., Behrens, T. E., Kennerley, S. W., & Rushworth, M. F. (2007). Adaptive decision making and value in the anterior cingulate cortex. *NeuroImage*, *36*(Suppl 2), T142–T154.
- Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual Review of Neuroscience*, *40*, 373–394.
- Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A., & Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron*, *74*, 858–873.
- Wei, W., & Wang, X. J. (2016). Inhibitory control in the cortico-basal ganglia-thalamocortical loop: Complex regulation and interplay with memory and decision processes. *Neuron*, *92*, 1093–1105.
- Weiss, C., & Disterhoft, J. F. (2015). The impact of hippocampal lesions on trace-eyeblick conditioning and forebrain-cerebellar interactions. *Behavioral Neuroscience*, *129*, 512–522.
- Wikenheiser, A. M., Marrero-Garcia, Y., & Schoenbaum, G. (2017). Suppression of ventral hippocampal output impairs integrated orbitofrontal encoding of task structure. *Neuron*, *95*, 1197–1207.e3.
- Wikenheiser, A. M., & Redish, A. D. (2015). Hippocampal theta sequences reflect current goals. *Nature Neuroscience*, *18*, 289–294.
- Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science (New York, N.Y.)*, *261*, 1055–1058.
- Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science (New York, N.Y.)*, *265*, 676–679.
- Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia*, *48*, 2339–2356.
- Wise, S. P. (2008). Forward frontal fields: Phylogeny and fundamental function. *Trends in Neurosciences*, *31*, 599–608.
- Wunderlich, K., Dayan, P., & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience*, *15*, 786–791.
- Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *The European Journal of Neuroscience*, *19*, 181–189.
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *The European Journal of Neuroscience*, *22*, 513–523.
- Young, J. J., & Shapiro, M. L. (2011). Dynamic coding of goal-directed paths by orbital prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*, 5989–6000.

**How to cite this article:** Rusu SI, Pennartz CMA. Learning, memory and consolidation mechanisms for behavioral control in hierarchically organized cortico-basal ganglia systems. *Hippocampus*. 2020;30:73–98. <https://doi.org/10.1002/hipo.23167>