



Published in final edited form as:

Trends Neurosci. 2020 September ; 43(9): 681–694. doi:10.1016/j.tins.2020.06.006.

Hypothalamic interactions with large-scale neural circuits underlying reinforcement learning and motivated behavior

Bruno B. Averbeck, Elisabeth A. Murray

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, 20892-4415

Abstract

Biological agents adapt behavior to support the survival needs of the individual and the species. In this review we outline anatomical, physiological, and computational processes that support reinforcement learning. We describe two circuits in the primate brain linked to specific aspects of learning and goal-directed behavior. The ventral circuit, which includes the amygdala, ventral medial prefrontal cortex, and ventral striatum, has substantial connectivity with the hypothalamus. The dorsal circuit, which includes inferior parietal cortex, dorsal lateral prefrontal cortex, and the dorsal striatum, has minimal connectivity with the hypothalamus. The hypothalamic connectivity suggests distinct roles for these circuits. We propose that the ventral circuit defines behavioral goals and the dorsal circuit orchestrates behavior to achieve those goals.

Keywords

Reinforcement learning; frontal-striatal circuits; prefrontal cortex; basal ganglia; striatum; amygdala; hypothalamus; devaluation; motivation

Drives, reinforcement and learning

Adaptive behavior, or learning, is critical to survival. Learning is usually studied relative to rewards and punishments, or positive and negative reinforcement. In reinforcement learning (RL), the goal is to learn to make choices that maximize rewards and minimize punishments. Framing learning from this simple perspective has been useful in both biology and artificial intelligence [1, 2]. Rewards and punishments are fundamental in biology because they drive learning which supports survival of the individual and the species. The mechanisms that have evolved to support learning have been selected in our ancestors because they provided an advantage for those agents, in their time and place, in some domain[3].

Theoretical accounts of RL assume rewards are a static property of the environment, for example a drop of juice or sucrose for a monkey or a rat. For complex organisms, like

Lead Contact: Bruno B. Averbeck, Ph.D., Laboratory of Neuropsychology, NIMH/NIH, Building 49 Room 1B80, 49 Convent Drive MSC 4415, Bethesda, MD 20892-4415, bruno.averbeck@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

humans, the definition of reward can appear arbitrary. Some people may find running a marathon to be rewarding, whereas others may find the same activity to be punishing, preferring instead to spend a relaxing day at home reading a book. At some level, however, a reward is relevant to a drive or a physiological need [4], although rewards do not have to be defined relative to an immediate drive or need [5]. Eating, drinking, sexual intercourse, rearing young, sleeping, avoiding predators, and sitting by a fire in the winter, are all associated directly with fundamental survival needs of the individual and the species, and each appears to be under the control of different hypothalamic circuits[6]. Other rewards, particularly in more complex species, may have more to do with maintaining position in a social hierarchy or obtaining secondary reinforcers like money, which in turn provide future access to resources that can be used to satisfy drives or needs. Because rewards are relevant to physiological needs, they are fundamentally dependent on an organism's internal states, and therefore rewards are not fixed properties of the environment[7, 8].

There is, therefore, a close relationship between drives, reinforcement, and learning. Although this has long been appreciated in learning theory, it is not often considered in biological studies of RL. Hull's drive reduction theory suggested that reinforcement followed from satisfaction of an internal need, and actions which satisfied internal needs were reinforced [4]. The earliest studies on intracranial self-stimulation, which are experiments in which animals will press a lever to receive electrical stimulation at specific locations in the brain, found that stimulation at a single site could lead to consummatory behavior, if for example food was available, as well as instrumental behavior to obtain more stimulation [9, 10]. These self-stimulation effects, which have been substantiated by modern methods, were particularly prevalent in the lateral hypothalamus [11, 12]. From these and other studies, several assumptions about learning have arisen: 1) motivation follows from elevated drives or needs; 2) satisfaction or predicted satisfaction of drives is reinforcing; and 3) learning systems operate to increase positive reinforcement and decrease negative reinforcement. More specifically, learning links objects in the environment with their relevance to specific drives or needs [13, 14]. This is the knowledge of what objects can do to or for us. Reinforcement and learning is, therefore, more sophisticated than just approach/avoid, positive/negative. Although the positive/negative dichotomy forms a useful experimental starting point, it obfuscates the true nature and complexity of learning. Much of the forebrain may be organized around learning to satisfy need-specific drives signaled by the hypothalamus. This is consistent with recent developmental models which suggest that the entire telencephalon is a massive dorsal expansion of the caudal hypothalamus [15].

In this review, we will consider progress in anatomy, physiology and theory which is beginning to define the neural architecture and computational mechanisms that underlie these behavioral processes. We will begin by outlining the large-scale anatomical organization of two brain circuits in primates, which we refer to as the dorsal and ventral circuits, and linking that circuitry to specific aspects of learning and goal-directed behavior. We will then discuss recent research which has defined drive-specific neural circuitry in the hypothalamus, and how interaction of the hypothalamus with the ventral circuit, including for example the ventral tegmental area (VTA), insular cortex, and amygdala, supports motivated behavior. In considering motivated behavior we focus on both reinforcer devaluation experiments, which examine need-specific adaptive behavior, and reinforcement

learning studies, which characterize learning to obtain reward and avoid punishment more generally. Finally, we consider these ideas relative to RL theory, and suggest that learning in biology is most often state-value, goal-directed learning, not action-value learning.

Anatomical organization of dorsal and ventral neural systems

An overarching framework for understanding the organization of visual processing has been the division of parietal and temporal cortex into dorsal and ventral streams [16]. The dorsal stream, which includes the inferior parietal cortex, contributes to spatial attention and perception, as well as to visually guided action. The ventral stream, which includes the inferior temporal cortex, has been characterized as subserving object identification and related processes [3]. Here, we outline prefrontal and subcortical extensions of these circuits. We will further suggest, based on the anatomy and known properties of nodes in the cortical-subcortical circuits, that the ventral circuitry identifies goals, and the dorsal circuitry computes actions flexibly to obtain those goals.

Frontal circuitry is organized in cortical-basal ganglia-thalamo-cortical loops [17, 18]. If we consider only the cortical-striatal component of this circuitry, there is a topographic organization, such that adjacent areas within prefrontal cortex project to adjacent areas within the striatum. There is considerable overlap in the terminal fields within the striatum of nearby prefrontal cortical areas [17, 19]. However, dorsal lateral prefrontal cortex (dlPFC) projects to the dorsal striatum (Fig. 1A, C), whereas ventral medial prefrontal cortex (vmPFC) projects to the ventral striatum (VS; Fig. 1B, C [18]). Therefore, widely separated prefrontal cortical areas have minimal overlap, and dlPFC and vmPFC have the most widely separated projections into the striatum [18, 19].

The segregated projections from cortex into the striatum continue through the pallidum and medial dorsal (MD) thalamus. The dlPFC projects to the dorsal striatum, which then projects to the globus pallidus internal segment (GPi); the GPi projects, in turn, to the lateral portion of the MD thalamus, which projects back to dlPFC (Fig. 1A, C). The vmPFC projects to the VS, which then projects to the ventral pallidum (VP [17, 20]); the VP projects, in turn, to the medial portion of the MD thalamus, which projects back to the vmPFC ([21] Fig. 1B, C).

This segregated circuitry reflects a broader large-scale organization of neural circuits within the primate brain. The dlPFC is strongly connected with area 7a in inferior parietal cortex (Fig. 1A). Area 7a also sends projections to the dorsal striatum and the lateral portion of the MD, that overlap with the dlPFC projections to the same structures [22] (Fig. 1A). The connected parietal-frontal circuit, therefore, has convergent projections to the striatum and thalamus. The vmPFC circuit has a parallel large-scale organization. The amygdala (and nearby medial temporal lobe cortex including entorhinal and perirhinal cortex [23]) projects to the vmPFC (Fig. 1B, Fig. 2; [24]) and also projects to the VS and the medial MD ([25, 26]). The connected temporal-frontal circuit, therefore, also has convergent projections to the striatum and thalamus. Thus, dorsal (Fig. 1A) and ventral (Fig. 1B) networks span parietal-prefrontal and medial temporal-prefrontal cortex. Each cortical circuit has convergent striatal and thalamic targets that, however, differ between networks.

There are interactions between the dorsal and ventral circuits, particularly within prefrontal cortex. However, connections tend to be stronger, even at the cortical level, *within* dorsal network nodes and *within* ventral network nodes, than *between* dorsal and ventral networks [27]. There are also cortical areas situated between dlPFC and vmPFC, including ventral lateral prefrontal cortex (vlPFC) and the anterior cingulate cortex (ACC; area 24). ACC and vlPFC have overlapping projections into the central striatum [18, 19]. Given that the ACC is anatomically connected to motor circuitry, and vlPFC is connected to sensory circuitry [28], the overlapping projections into the striatum may provide a site for linking actions and objects. More work will be necessary to understand this circuitry, however.

An important feature of the anatomical organization of the dorsal and ventral systems, and perhaps the most important feature from a functional perspective, is that the ventral systems, both the amygdala and vmPFC circuits, are connected strongly with the hypothalamus [6, 29] (Fig. 1C, Fig. 2), whereas the dorsal system has minimal direct connections to the hypothalamus [29]. There are strong projections from vmPFC, as well as the VS and VP to the hypothalamus (Fig. 2). The CeN and the BNST also project to the hypothalamus [6]. The ventral system, therefore, is well poised to relay highly processed visual and other sensory information to the hypothalamus. The dorsal system, via the dorsal striatum, projects to the substantia nigra, which supports exploratory behavior. It has been proposed that this is a dorsal analog of the ventral circuit's projection to the hypothalamus proper [30].

Given the role of the hypothalamus in need-specific drives and motivated behavior (discussed below), this suggests a dissociation of function between the dorsal and ventral networks. Stated broadly, the ventral system defines behavioral goals, and the dorsal system orchestrates behavior to achieve those goals [3, 14, 31]. More specifically, the ventral system, by virtue of inputs from expanded inferior temporal lobe visual areas in primates and connections to the hypothalamus, is specialized for identification of objects and nonobject signs and features in the environment, for example auditory cues, that can satisfy needs or drives. For example, the ventral system can represent the current value (or desirability) and availability of objects which deliver specific nutrients [32]. The dorsal system, by virtue of inputs from parietal cortical areas, is specialized for representing spatial metrics related to actions undertaken to acquire goal objects. These metrics include the relative number of objects and distances between them, their speeds, and temporal quantities of actions including duration [31, 33].

With respect to learning, the ventral system learns the motivational value of objects in the environment that can satisfy specific needs [2]. It does this by combining information about which needs are satisfied, represented in the hypothalamus [6], and the objects that have led to that satisfaction, represented by inferior temporal visual inputs relayed through the medial temporal lobe structures. There may be less learning in the dorsal system, as the metrics for actions are relatively immutable ([34] see Outstanding Questions). The dorsal system may, however, learn hierarchically organized behaviors and habits, i.e. how to orchestrate behavior to achieve goals [35].

Drive-specific and motivational hypothalamic circuitry

The hypothalamus arguably supports all functions necessary for the survival of the individual and the species, although it does not, by itself, have the ability to learn novel ways to satisfy those needs. Animals with forebrain lesions rostral to, but not caudal to, the hypothalamus can maintain physiological homeostasis by ingesting food and water, and they also exhibit basic reproduction and defensive behaviors [30, 36]. Understanding the detailed circuitry of the hypothalamus has, however, been difficult because it is a complex structure composed of multiple nuclei buried deep in the brain. Further adding to the complexity of the hypothalamus is the cell-type and anatomical complexity within individual nuclei, many of which support multiple functions. Recent advances, using genetically identified cell types and circuit specific manipulations in mice, have led to a deeper understanding of the organization and function of circuit elements underlying multiple drive-specific homeostatic processes, including eating and drinking (Box 1). These studies in mice have shown that need-specific drive mechanisms can motivate behavior and support learning. Furthermore, the hypothalamus is conserved across the vertebrate lineage [15] and may have existed in a primordial form in the common ancestor to vertebrates and chordates [37]. Therefore, many of the findings in mice likely generalize to primates in which we have described the anatomical circuits, as well as other vertebrates.

Aside from the drive-specific hypothalamic systems (Box 1), the lateral hypothalamus (LH) may play a more general role in motivated behavior and reinforcement. Classical studies on the LH found that it supported intracranial self-stimulation and promoted feeding [9]. Modern studies have begun dissecting the specific circuit elements within the LH that support these behaviors. An optogenetics study in mice, for instance, has found that activation of GABAergic neurons in the LH (LH^{GABA}) increased food consumption and optical self-stimulation behavior [12]. These neurons were also active when mice nose-poked to obtain food reward, or consumed food reward [12]. A subsequent study examined specific projections from LH^{GABA} neurons to the VTA [38], as the LH provides a large input to the VTA. This study found that LH^{GABA}-VTA projections contributed to appetitive learning measured with both conditioned place preference and nose-pokes for optical activation of LH^{GABA} terminals in the VTA. These effects were primarily mediated by LH^{GABA} inputs to GABAergic interneurons within the VTA. The mechanism is thought to be through the LH inputs acting to disinhibit dopamine (DA) neurons, which leads to dopamine release in the VS. These results are consistent with the results from the earlier self-stimulation studies, which suggested that activation of the LH leads to feeding and is also rewarding. Other studies have found effects of activation of the LH^{GABA}-VTA circuit that depend on the stimulation frequency. Feeding was most strongly supported by low frequency optical stimulation (5 Hz) whereas optical self-stimulation was most strongly supported by higher frequency stimulation (40 Hz)[10].

In contrast to the LH^{GABA} neurons, glutamate neurons in the LH (LH^{Glut}) can mediate aversive learning. Specifically, activation of LH^{Glut}-VTA projections leads to a decrease in VTA dopamine release [38]. Subsequent experiments which carried out a more detailed examination of LH^{Glut}-VTA effects on dopamine release in the VS found a circuit from the VTA to the medial VS shell that supported aversive learning [39]. Activation of LH^{Glut}-VTA

preferentially leads to increased dopamine release in the medial shell, decreased dopamine release in other VS regions, and real-time place aversion. Furthermore, inactivating the LH^{Glut}-VTA circuit led to decreased avoidance of a noxious odor. Additional studies have found that LH^{Glut} neurons also project to the lateral habenula (LHb). Strengthening the LH^{Glut}-LHb synaptic connectivity underlies components of avoidance learning [40] and activation of LH^{Glut}-LHb neurons (but not GPi^{Glut}-LHb neurons) also induces avoidance learning [41], and LH^{Glut} neurons respond to, and come to predict, aversive events [42]. In addition, inhibiting the LH^{Glut}-LHb circuit impairs aversive learning [43]. Given the indirect connections of the LHb and VTA via the rostral-medial tegmental nucleus [44], the LH^{Glut}-LHb circuit provides another route by which aversive information can reach DA neurons. In summary, these studies suggest that the LH^{GABA}-VTA circuit can support appetitive learning by disinhibiting VTA GABAergic local circuit neurons, and therefore disinhibiting DA neurons. The LH^{Glut} neurons, by contrast, support aversive learning by projecting to a specific sub-population of VTA neurons that projects to the medial shell of the VS, as well as projections to the LHb. This circuitry can likely support need-specific learning as well, but to our knowledge, this has not yet been studied.

Role of internal state in adaptive behavior and learning - motivation and devaluation

The hypothalamus can support drive-specific motivated behavior which can drive learning, but the hypothalamus and the associated ventral circuitry also contribute more broadly to adaptive behavior. Implicit in studies of learning is the fact that animals learn because they are motivated, usually by hunger or thirst in laboratory settings, and the experiments are designed so the subjects can learn to satisfy those drives. After animals have consumed substantial food or water, they no longer carry out tasks. Anecdotally, anyone who has ever studied behavior in the laboratory knows this to be true; animals need to be motivated to perform tasks. In one example, in experiments in which a simple motor response led to rewards of different sizes or at different delays, with reward size and delay explicitly indicated by a visual cue at the start of the trial, monkeys aborted trials in which smaller rewards or longer delays were offered. These effects, however, depended on the level of satiation [45, 46]. When animals were more satiated, later in the session, they aborted significantly more trials than they did early in the session. Therefore, the relevant behavioral state for these animals included not only the available choice options, but also a representation of the monkey's current needs or drives. These motivational processes were disrupted by lesions that disrupted the direct interaction between the orbitofrontal cortex and the perirhinal and entorhinal cortex [46]. This supports a role for the perirhinal and entorhinal areas in relaying visual cue information, which in these experiments defines reward size and delay, to the broader ventral circuitry.

Both the insular cortex and the amygdala play important roles in interfacing between hypothalamic drive signals, internal states and external (environmental) stimuli. Neurons in the insular cortex of rodents encode levels of thirst and hunger in largely nonoverlapping populations of neurons. Thus, insular cortex neurons represent the physiological state of an animal, at least with respect to hunger or thirst [47]. Related work has shown that cues that

predict food delivery in behavioral tasks activate insular cortex neurons when animals are deprived, but not when they are satiated [48]. In these studies, mice were trained to lick following presentation of a cue which predicted food delivery and to withhold licking in response to a different cue which predicted aversive quinine delivery. Neurons in insular cortex responded more to presentation of the cue when the animals were deprived than satiated. Notably, visual predictive cues and subsequent consumption of a small amount of water or food in thirsty or hungry mice shifted the activity patterns in insular cortex to the pattern observed during the future satiated state, regardless of the current level of food or water deficit. Thus, in response to environmental cues, insular cortex neuronal activity anticipates, or predicts, future internal states. Additionally, neurons in insular cortex responded to the cue in satiated animals following activation of Arc^{AgRP} neurons, activation of which drives feeding behavior (Box 1). The authors found a multi-synaptic circuit from Arc^{AgRP} to PVT to the BLA and finally to insular cortex that mediated these effects, as well as related effects driven by thirst [47]. Future studies should aim to identify analogous circuitry in human and nonhuman primates.

The amygdala interacts with the hypothalamus to mediate the influence of learned cues on behavior. Although these mechanisms normally aid animals in maintaining homeostasis—by directing them to cues that reliably predict reduction of drives—they can also operate in a counterintuitive manner. The amygdala can serve to increase eating in satiated rats and to stop eating in hungry rats [49]. For example, when animals are in the presence of cues that predict food, interaction of the basolateral amygdala with the lateral hypothalamus overrides satiety and promotes eating [50], consistent with insensitive salience[14].

The ventral circuits also influence choice preference in hungry animals. One fruitful avenue of research to examine this employs the devaluation task. In the version of the task typically used with monkeys, food value is manipulated using a selective satiation procedure intended to devalue the food, and the effect of that manipulation is measured as a shift in both food preference and object choices. Monkeys are first trained to associate objects with specific foods. Some objects are associated with one type of food (for example peanuts) and other objects associated with a different food (for example chocolate chips). In a series of probe tests, monkeys are—for the first time—offered a choice between two rewarded objects: a peanut-associated object and a chocolate-chip associated object. There is no wrong answer because each of the objects overlies the food with which it was paired in the training phase. The probe tests measure the monkey's choices in three conditions tested on separate days: baseline choices, choices of objects after being fed to satiety on one food (for example peanuts), and choices after being fed to satiety on the other food (for example chocolate chips) [51]. After being satiated on one of the two foods, intact monkeys show a preference for the objects overlying the nonsatiated food, which is instantiated as a shift in choices of objects, termed the devaluation effect. Lesions of the amygdala[52], orbitofrontal cortex[53], ventral striatum [54], or the medial MD thalamus[55] lead to a loss of devaluation effects. Animals with lesions to these structures and unoperated controls alike show a strong preference for the nonsatiated food when given either consumption tests or two-alternative forced-choices between the satiated and nonsatiated foods. Animals in the lesion groups, however, no longer show a preference for the objects that predict foods that they have not recently eaten, relative to those that they have. Thus, lesions across multiple nodes of the

ventral network disrupt devaluation effects. Importantly, the same lesions do not disrupt visual sensory processing, willingness to work for foods, or general satiety mechanisms. Thus the lesions selectively disrupt the ability of animals to link objects with the sensory properties, including the current value, of the foods they predict. Motivation drives learning and we learn to make choices that satisfy specific needs. The devaluation studies point to a role for the amygdala and orbitofrontal cortex, together with other regions, in linking environmental cues with specific foods, thereby linking the need-specific drives of the hypothalamus with the frontal cortex areas that select goals for action.

Devaluation and related behaviors show that motivation and preference are adapted in response to changes in needs. The same circuitry, however, motivates animals to learn to satisfy needs. Learning systems in the brain operate to increase positive reinforcement and decrease negative reinforcement. Reinforcement learning (RL; Box 2) provides a theoretical framework for understanding links between drives, largely signaled by the hypothalamus (Box 1), and learning systems, which acquire and store knowledge of which objects and actions can satisfy drives. How do the component processes of RL, including states, s_t , state values, $V(s_t)$, and action values, $Q(s_t, a_t)$, map onto neural circuitry (Figs. 1 and 2)? States are fundamental to RL, because they are the information on which choices (policies) and learning are based [56]. Therefore, it is important to understand how and where states are represented in the brain. Motivated behavior depends on both internal needs and the availability of relevant external objects [5, 14, 50]. It is clear from classic and more recent work that activity within the specific drive systems in the hypothalamus (Box 1) can motivate behavior and support learning by representing need-specific internal states, $s_{internal}$ [57]. Information about objects in the external environment, $s_{external}$, are relayed to the ventral circuitry (Fig. 1B, Fig. 2) by the amygdala, entorhinal and perirhinal cortex [58–60]. These cortical nodes define the entry points for visual information into the ventral circuitry, along with the hippocampal region (Box 3). Therefore, specification of the full state, s_t , necessary for motivation and learning has to take place across nodes within the ventral circuitry, with $s_{internal}$ being relayed from the hypothalamus and brainstem, and $s_{external}$ mostly visual information in primates, entering via the medial temporal lobe cortical nodes.

In addition to states, the values of states, $V(s_t)$, and the values of actions available in those states, $Q(s_t, a_t)$, have to be learned and represented. State value learning (Box 2) is carried out in the ventral circuitry [58, 60–62]. Physiological recordings across nodes in the ventral circuitry have shown that neurons in this pathway code information about objects, the associated values of those objects, and relevant internal states [58, 60, 63, 64]. Therefore, they code both internal and external states, s_t , and state values $V(s_t)$. The ventral circuitry, however, has minimal information about specific actions [58, 63]. Action values, $Q(s_t, a_t)$, are encoded in the dorsal circuitry, including the dorsal striatum [35, 65]. More broadly, neurons in the dorsal circuit encode information about the actions necessary to acquire goal objects [66–68]. Although these studies on the dorsal circuitry appear to support a role for the dorsal circuitry in learning and representing action values, they are based on a behaviorist S-R conception of action value, where a_t refers to a specific action, e.g. saccade left or press the left lever. This follows from the operant paradigms often used in laboratory experiments. These experiments constrain goals and therefore reinforcement to specific actions. It remains

unclear, then, whether action values or goal states are being encoded in dorsal circuitry, because the goals are confounded with the action themselves (see Outstanding Questions).

Although action values, $Q(s_b, a_t)$, are fundamental to theoretical accounts of RL, they may be less relevant than state values, $V(s_t)$, in biology. We would argue that given the limited connectivity of the dorsal circuit with the hypothalamus, it is unlikely that the dorsal circuit, and therefore action values represented in the dorsal circuit, motivate behavior or define the goals of motivated behavior (see Outstanding Questions). (Although, it should be noted, habitual action sequences are important, and these could be represented by action value learning in dorsal circuitry[35].) In state value learning, the values of specific actions or action sequences are not learned. Rather, state values that specify goals relevant to specific needs are learned. Goals are future target states, s_{t+k} (where $t+k$ indicates a future time) in which an immediate reward, $r(s_b, a_t)$, is received. After goal objects are identified in the ventral circuitry, the dorsal circuitry calculates relevant actions, on the fly, to obtain the goal state [14]. For example, in predator-prey interactions, the goals are clear, and the actions are computed flexibly, moment-to-moment, to achieve those goals. Consistent with this, dorsal circuit neurons encode not only actions, but also goal objects and their values [66, 68]. However, the values of goal objects are likely inherited from the ventral circuit [62, 69, 70]. Intermediate areas, like vIPFC or the ACC may play a particularly important role in translating state values into relevant actions [71]. Given that bilateral lesions of vIPFC fail to disrupt performance on devaluation tasks, however, the vIPFC cannot be considered an obligatory route for converting goals into the relevant actions[52, 72].

Concluding Remarks

In this review we have proposed a behavioral, anatomical and computational framework for conceptualizing adaptive, motivated behavior. We have suggested that motivated behavior follows from physiological needs, or predicted future needs (which are at least partially mediated by circadian processes[57]). Satisfaction or predicted satisfaction of these needs is reinforcing, and learning systems adapt behavior to increase positive reinforcement and decrease negative reinforcement, which maximizes need satisfaction. We further propose that cortical and subcortical neural systems, at least in primates, are organized into two large-scale networks that underlie these processes. The dorsal network is specialized for representing spatial metrics to achieve goals, including relative distances, speed, number, and duration; and computing appropriate actions flexibly based on those metrics. The ventral network, via its interaction with the hypothalamus, mediates learning about state values, which identify objects, as well as signs and features in the environment that can satisfy current or future needs. Furthermore, subjects associate objects in the environment with satisfaction of specific drives, and those links are formed through learning. Subjects learn about what the objects in the environment can do to or for them. Learning, therefore, is more sophisticated than the approach avoid dichotomy. Need-specific associations are currently poorly understood, beyond the coarse classification of motivation into positive and negative reinforcement. Need-specific learning has also not been typically incorporated into RL theory applied in biology (see Outstanding Questions), although the RL framework can readily capture these effects (Box 2). Finally, we have described the dorsal and ventral networks and outlined their role in satisfying immediate drives. Further empirical work will be

necessary to understand the interaction between these networks and the frontal pole (area 10) and anterior prefrontal networks that mediate abstract planning and motivated pursuit of long-term goals [73–75].

Acknowledgements

This work was supported by the Intramural Research Program of NIMH (BA: ZIA MH002928, EM: ZIA MH002887).

References

1. Neftci EO and Averbeck BB (2019) Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence* 1, 133–143.
2. Averbeck BB and Costa VD (2017) Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience* 20(4), 505–512. [PubMed: 28352111]
3. Murray EA et al. (2017) *The evolution of memory systems: ancestors, anatomy, and adaptations*, First Edition edn., Oxford University Press.
4. Hull CL (1943) *Principles of behavior, an introduction to behavior theory*, D. Appleton-Century Company.
5. Toates FM (1986) *Motivational systems*, Cambridge University Press.
6. Risold PY et al. (1997) The structural organization of connections between hypothalamus and cerebral cortex. *Brain Res Brain Res Rev* 24(2–3), 197–254. [PubMed: 9385455]
7. van Swieten MMH and Bogacz R (2020) Modeling the effects of motivation on choice and learning in the basal ganglia. *PLoS Comput Biol* 16(5), e1007465. [PubMed: 32453725]
8. Cone JJ et al. (2016) Physiological state gates acquisition and expression of mesolimbic reward prediction signals. *Proc Natl Acad Sci U S A* 113(7), 1943–8. [PubMed: 26831116]
9. Olds J (1977) *Drives and Reinforcements: Behavioral studies of hypothalamic functions*, Raven Press.
10. Barbano MF et al. (2016) Feeding and Reward Are Differentially Induced by Activating GABAergic Lateral Hypothalamic Projections to VTA. *J Neurosci* 36(10), 2975–85. [PubMed: 26961951]
11. Olds J et al. (1960) Topographic Organization of Hypothalamic Self-Stimulation Functions. *Journal of Comparative and Physiological Psychology* 53(1), 23–32. [PubMed: 14428621]
12. Jennings JH et al. (2015) Visualizing hypothalamic network dynamics for appetitive and consummatory behaviors. *Cell* 160(3), 516–27. [PubMed: 25635459]
13. Burgess CR et al. (2018) Gating of visual processing by physiological need. *Curr Opin Neurobiol* 49, 16–23. [PubMed: 29125986]
14. Bindra D (1978) How Adaptive-Behavior Is Produced - Perceptual-Motivational Alternative to Response-Reinforcement. *Behavioral and Brain Sciences* 1(1), 41–52.
15. Puelles L and Rubenstein JL (2015) A new scenario of hypothalamic organization: rationale of new hypotheses introduced in the updated prosomeric model. *Front Neuroanat* 9, 27. [PubMed: 25852489]
16. Ungerleider LG and Mishkin M (1982) Two cortical visual systems. In *Analysis of Visual Behavior* (Ingle DJ et al. eds), pp. 549–587, MIT Press.
17. Alexander GE et al. (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9, 357–81. [PubMed: 3085570]
18. Haber SN et al. (2006) Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J Neurosci* 26(32), 8368–76. [PubMed: 16899732]
19. Averbeck BB et al. (2014) Estimates of projection overlap and zones of convergence within frontal-striatal circuits. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 34(29), 9497–505. [PubMed: 25031393]

20. Root DH et al. (2015) The ventral pallidum: Subregion-specific functional anatomy and roles in motivated behaviors. *Prog Neurobiol* 130, 29–70. [PubMed: 25857550]
21. Giguere M and Goldman-Rakic PS (1988) Mediodorsal nucleus: areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. *J Comp Neurol* 277(2), 195–213. [PubMed: 2466057]
22. Selemon LD and Goldman-Rakic PS (1988) Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *J Neurosci* 8(11), 4049–68. [PubMed: 2846794]
23. Munoz M and Insausti R (2005) Cortical efferents of the entorhinal cortex and the adjacent parahippocampal region in the monkey (*Macaca fascicularis*). *Eur J Neurosci* 22(6), 1368–88. [PubMed: 16190892]
24. Ghashghaei HT et al. (2007) Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage* 34(3), 905–23. [PubMed: 17126037]
25. Friedman DP et al. (2002) Comparison of hippocampal, amygdala, and perirhinal projections to the nucleus accumbens: combined anterograde and retrograde tracing study in the Macaque brain. *The Journal of Comparative Neurology* 450(4), 345–65. [PubMed: 12209848]
26. Russchen FT et al. (1987) The afferent input to the magnocellular division of the mediodorsal thalamic nucleus in the monkey, *Macaca fascicularis*. *J Comp Neurol* 256(2), 175–210. [PubMed: 3549796]
27. Barbas H and Pandya DN (1989) Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 286(3), 353–75. [PubMed: 2768563]
28. Averbeck BB and Seo M (2008) The statistical neuroanatomy of frontal networks in the macaque. *PLoS Comput Biol* 4(4), e1000050. [PubMed: 18389057]
29. Rempel-Clower NL and Barbas H (1998) Topographic organization of connections between the hypothalamus and prefrontal cortex in the rhesus monkey. *J Comp Neurol* 398(3), 393–419. [PubMed: 9714151]
30. Swanson LW (2000) Cerebral hemisphere regulation of motivated behavior. *Brain Res* 886(1–2), 113–164. [PubMed: 11119693]
31. Andersen RA et al. (1997) Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu Rev Neurosci* 20, 303–30. [PubMed: 9056716]
32. Murray EA and Rudebeck PH (2018) Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. *Nat Rev Neurosci* 19(7), 404–417. [PubMed: 29795133]
33. Genovesio A et al. (2014) Prefrontal-parietal function: from foraging to foresight. *Trends Cogn Sci* 18(2), 72–81. [PubMed: 24378542]
34. Garcia-Cabezas MA et al. (2019) The Structural Model: a theory linking connections, plasticity, pathology, development and evolution of the cerebral cortex. *Brain Struct Funct* 224(3), 985–1008. [PubMed: 30739157]
35. Seo M et al. (2012) Action selection and action value in frontal-striatal circuits. *Neuron* 74(5), 947–60. [PubMed: 22681697]
36. Hinsey JC et al. (1930) The role of the hypothalamus and mesencephalon in locomotion. *Archives of Neurology and Psychiatry* 23(1), 1–43.
37. Lacalli T (2018) Amphioxus neurocircuits, enhanced arousal, and the origin of vertebrate consciousness. *Conscious Cogn* 62, 127–134. [PubMed: 29598920]
38. Nieh EH et al. (2016) Inhibitory Input from the Lateral Hypothalamus to the Ventral Tegmental Area Disinhibits Dopamine Neurons and Promotes Behavioral Activation. *Neuron* 90(6), 1286–1298. [PubMed: 27238864]
39. de Jong JW et al. (2019) A Neural Circuit Mechanism for Encoding Aversive Stimuli in the Mesolimbic Dopamine System. *Neuron* 101(1), 133–151 e7. [PubMed: 30503173]
40. Trusel M et al. (2019) Punishment-Predictive Cues Guide Avoidance through Potentiation of Hypothalamus-to-Habenula Synapses. *Neuron* 102(1), 120–127 e4. [PubMed: 30765165]

41. Stamatakis AM et al. (2016) Lateral Hypothalamic Area Glutamatergic Neurons and Their Projections to the Lateral Habenula Regulate Feeding and Reward. *J Neurosci* 36(2), 302–11. [PubMed: 26758824]
42. Lazaridis I et al. (2019) A hypothalamus-habenula circuit controls aversion. *Mol Psychiatry* 24(9), 1351–1368. [PubMed: 30755721]
43. Lecca S et al. (2017) Aversive stimuli drive hypothalamus-to-habenula excitation to promote escape behavior. *Elife* 6.
44. Lammel S et al. (2012) Input-specific control of reward and aversion in the ventral tegmental area. *Nature* 491(7423), 212–7. [PubMed: 23064228]
45. Minamimoto T et al. (2009) Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys. *J Neurophysiol* 101(1), 437–47. [PubMed: 18987119]
46. Clark AM et al. (2013) Interaction between orbital prefrontal and rhinal cortex is required for normal estimates of expected value. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 33(5), 1833–45. [PubMed: 23365223]
47. Livneh Y et al. (2020) Estimation of Current and Future Physiological States in Insular Cortex. *Neuron*.
48. Livneh Y et al. (2017) Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature* 546(7660), 611–616. [PubMed: 28614299]
49. Petrovich GD (2013) Forebrain networks and the control of feeding by environmental learned cues. *Physiol Behav* 121, 10–8. [PubMed: 23562305]
50. Petrovich GD et al. (2002) Amygdalo-hypothalamic circuit allows learned cues to override satiety and promote eating. *J Neurosci* 22(19), 8748–53. [PubMed: 12351750]
51. Murray EA and Rudebeck PH (2013) The drive to strive: goal generation based on current needs. *Frontiers in neuroscience* 7, 112. [PubMed: 23818876]
52. Murray EA and Rudebeck PH (2018) Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. *Nature Reviews Neuroscience*.
53. Murray EA et al. (2015) Specialized areas for value updating and goal selection in the primate orbitofrontal cortex. *Elife* 4.
54. Singh T et al. (2010) Nucleus Accumbens Core and Shell are Necessary for Reinforcer Devaluation Effects on Pavlovian Conditioned Responding. *Frontiers in integrative neuroscience* 4, 126. [PubMed: 21088698]
55. Mitchell AS (2015) The mediodorsal thalamus as a higher order thalamic relay nucleus important for learning and decision-making. *Neurosci Biobehav Rev* 54, 76–88. [PubMed: 25757689]
56. Averbeck BB (2015) Theory of choice in bandit, information sampling and foraging tasks. *PLoS computational biology* 11(3), e1004164. [PubMed: 25815510]
57. Challet E (2019) The circadian regulation of food intake. *Nat Rev Endocrinol* 15(7), 393–405. [PubMed: 31073218]
58. Costa VD et al. (2019) Subcortical Substrates of Explore-Exploit Decisions in Primates. *Neuron* 103 (3), 533–545 e5. [PubMed: 31196672]
59. Rudebeck PH et al. (2013) Effects of amygdala lesions on reward-value coding in orbital and medial prefrontal cortex. *Neuron* 80(6), 1519–31. [PubMed: 24360550]
60. Rudebeck PH et al. (2017) Amygdala Contributions to Stimulus-Reward Encoding in the Macaque Medial and Orbital Frontal Cortex during Learning. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 37(8), 2186–2202. [PubMed: 28123082]
61. Belova MA et al. (2008) Moment-to-moment tracking of state value in the amygdala. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 28(40), 10023–30. [PubMed: 18829960]
62. Costa VD et al. (2016) Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning. *Neuron* 92(2), 505–517. [PubMed: 27720488]
63. Costa VD and Averbeck BB (2020) Primate orbitofrontal cortex codes information relevant for managing explore-exploit tradeoffs. *J Neurosci*.

64. Bouret S and Richmond BJ (2010) Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *J Neurosci* 30(25), 8591–601. [PubMed: 20573905]
65. Lee E et al. (2015) Injection of a Dopamine Type 2 Receptor Antagonist into the Dorsal Striatum Disrupts Choices Driven by Previous Outcomes, But Not Perceptual Inference. *Journal of Neuroscience*.
66. Wallis JD and Miller EK (2003) Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *The European journal of neuroscience* 18(7), 2069–81. [PubMed: 14622240]
67. Bartolo R and Averbeck BB (2020) Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron*.
68. Bartolo R et al. (2020) Dimensionality, information and learning in prefrontal cortex. *PLoS Comput Biol* 16(4), e1007514. [PubMed: 32330126]
69. Rothenhoefer KM et al. (2017) Effects of ventral striatum lesions on stimulus versus action based reinforcement learning. *Journal of Neuroscience*.
70. Taswell CA et al. (2018) Ventral striatum's role in learning from gains and losses. *Proc Natl Acad Sci U S A* 115(52), E12398–E12406. [PubMed: 30545910]
71. Cai X and Padoa-Schioppa C (2014) Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* 81(5), 1140–1151. [PubMed: 24529981]
72. Rudebeck PH et al. (2017) Specialized Representations of Value in the Orbital and Ventrolateral Prefrontal Cortex: Desirability versus Availability of Outcomes. *Neuron* 95(5), 1208–1220 e5. [PubMed: 28858621]
73. Badre D and D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci* 10(9), 659–69. [PubMed: 19672274]
74. Dixon ML et al. (2017) Hierarchical Organization of Frontoparietal Control Networks Underlying Goal-Directed Behavior. In *The Prefrontal Cortex as an Executive, Emotional, and Social Brain* (Watanabe M ed), Springer.
75. Tsujimoto S et al. (2011) Frontal pole cortex: encoding ends at the end of the endbrain. *Trends Cogn Sci* 15(4), 169–76. [PubMed: 21388858]
76. Sternson SM and Eiselt AK (2017) Three Pillars for the Neural Control of Appetite. *Annu Rev Physiol* 79, 401–423. [PubMed: 27912679]
77. Andermann ML and Lowell BB (2017) Toward a Wiring Diagram Understanding of Appetite Control. *Neuron* 95(4), 757–778. [PubMed: 28817798]
78. Chen Y et al. (2016) Hunger neurons drive feeding through a sustained, positive reinforcement signal. *Elife* 5.
79. Betley JN et al. (2013) Parallel, redundant circuit organization for homeostatic control of feeding behavior. *Cell* 155(6), 1337–50. [PubMed: 24315102]
80. Betley JN et al. (2015) Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature* 521(7551), 180–185. [PubMed: 25915020]
81. Allen WE et al. (2017) Thirst-associated preoptic neurons encode an aversive motivational drive. *Science* 357(6356), 1149–1155. [PubMed: 28912243]
82. Krashes MJ et al. (2011) Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J Clin Invest* 121 (4), 1424–8. [PubMed: 21364278]
83. Leib DE et al. (2017) The Forebrain Thirst Circuit Drives Drinking through Negative Reinforcement. *Neuron* 96(6), 1272–1281 e4. [PubMed: 29268095]
84. Berridge KC (2004) Motivation concepts in behavioral neuroscience. *Physiol Behav* 81(2), 179–209. [PubMed: 15159167]
85. Miller NE and Kessen ML (1952) Reward effects of food via stomach fistula compared with those of food via mouth. *J Comp Physiol Psychol* 45(6), 555–64. [PubMed: 13000028]
86. Myers KP and Hall WG (1998) Evidence that oral and nutrient reinforcers differentially condition appetitive and consummatory responses to flavors. *Physiol Behav* 64(4), 493–500. [PubMed: 9761223]

87. Mcfarland D (1969) Separation of Satiating and Rewarding Consequences of Drinking. *Physiology & Behavior* 4(6), 987-+.
88. Augustine V et al. (2019) Temporally and Spatially Distinct Thirst Satiation Signals. *Neuron* 103(2), 242–249 e4. [PubMed: 31153646]
89. Collins AG and Frank MJ (2014) Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review* 121(3), 337–66. [PubMed: 25090423]
90. Colas JT et al. (2017) Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI. *PLoS Comput Biol* 13(10), e1005810. [PubMed: 29049406]
91. Takahashi Y et al. (2008) Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Front Neurosci* 2(1), 86–99. [PubMed: 18982111]
92. Botvinick M and Weinstein A (2014) Model-based hierarchical reinforcement learning and human action control. *Philos Trans R Soc Lond B Biol Sci* 369(1655).
93. Franklin NT and Frank MJ (2020) Generalizing to generalize: Humans flexibly switch between compositional and conjunctive structures during reinforcement learning. *PLoS Comput Biol* 16(4), e1007720. [PubMed: 32282795]
94. Kravitz DJ et al. (2011) A new neural framework for visuospatial processing. *Nat Rev Neurosci* 12(4), 217–30. [PubMed: 21415848]
95. Aggleton JP et al. (2015) Complementary Patterns of Direct Amygdala and Hippocampal Projections to the Macaque Prefrontal Cortex. *Cereb Cortex* 25(11), 4351–73. [PubMed: 25715284]
96. Averbeck BB (2019) Pavlovian patterns in the amygdala. *Nat Neurosci* 22(12), 1949–1950. [PubMed: 31768053]
97. Gershman SJ and Daw ND (2017) Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology* 68, 101–128.
98. Hsu DT and Price JL (2009) Paraventricular thalamic nucleus: subcortical connections and innervation by serotonin, orexin, and corticotropin-releasing hormone in macaque monkeys. *J Comp Neurol* 512(6), 825–48. [PubMed: 19085970]

Box 1:**Need-specific hypothalamic circuitry**

Recently, there have been substantial advances in understanding the hypothalamic circuitry underlying hunger and thirst (reviewed in detail elsewhere [76, 77]), based on studies in mice. Neurons important for feeding include agouti-related protein (AgRP) and pro-opiomelanocortin (POMC) expressing neurons in the arcuate nucleus of the hypothalamus (Arc). The Arc^{AgRP} and Arc^{POMC} neurons project to the paraventricular nucleus of the hypothalamus (PVH). The PVH neurons then project to the lateral parabrachial nucleus, which is also an important node in feeding. The Arc^{AgRP} neurons also project to the PVT, BNST and lateral hypothalamus (LH) and optogenetic activation of terminals from Arc^{AgRP} neurons in any of these structures drives feeding [78, 79]. Although activation of Arc^{AgRP} neurons leads to feeding, the activity of these neurons does not directly reflect homeostatic state. When mice are food restricted, Arc^{AgRP} neurons increase activity. However, the activity of Arc^{AgRP} neurons drops as soon as the animals are shown a chow pellet, before consumption [80]. They also transiently decrease activity when mice are presented with a visual cue that predicts delivery of a pellet [48]. Therefore, even at the level of the hypothalamus, the circuitry may have more to do with prediction of future states than reflection of current homeostatic needs [47].

In parallel with the feeding circuit, a neural circuit composed of the sub-fornical organ (SFO), the organum vasculosum of the lamina terminalis (OVLT) and the median preoptic nucleus (MnPO) is important for driving thirst. Both the SFO and the OVLT have neurons which detect changes in blood osmolality and the SFO also detects changes in angiotensin II. These neurons send excitatory projections to the MnPO, which then sends excitatory projections to the hypothalamus (PVH, LH and PVT) that drive thirst [81].

The circuitry controlling feeding and thirst can motivate animals to work. Animals will work for access to food following activation of Arc^{AgRP} neurons [82], and they will work for water following activation of SFO or thirst-related MnPO neurons [81]. Experiments have also shown that activating Arc^{AgRP} or thirst related MnPO neurons is aversive, and inactivation of these neurons can be reinforcing. Specifically, animals show conditioned place aversion to activation of Arc^{AgRP} neurons [80]. Interestingly, however, in the presence of food, animals will lever-press for activation of Arc^{AgRP} neurons [78]. In these experiments the animals engaged in lever-pressing followed by active feeding. Therefore, activation of Arc^{AgRP} neurons appears to be aversive in the absence of food and reinforcing when animals can activate the neurons and then feed. Activation of SFO^{NOS1} or identified thirst-related MnPO neurons also led to conditioned place aversion [80, 81]. Furthermore, animals learned to press a lever to turn off stimulation of SFO, OVLT or MnPO thirst-related neurons [83]. These neurons can, therefore, support need-specific drives, and they may do so by generating an aversive state.

These experiments support the drive-reduction hypothesis, which suggests that motivation follows from the need to reduce homeostatic drives [4, 84]. However, other work, using intragastric delivery of nutrition or fluid, has shown that motivation is related

to drive reduction in complex ways. When animals are provided with food or water intragastrically, they will continue to drink or eat [85–87]. Related to this, recent work has shown that orally delivered water leads to dopamine release in the striatum, whereas intragastric water delivery does not [88]. Both forms of delivery, however, lead to a decrease in activity of thirst related SFO neurons, although intragastric delivery decreased activity of SFO neurons more slowly. Therefore, oral-pharyngeal signaling is an important component of dopamine release, which likely is reinforcing, and drive reduction may depend on multiple factors including immediate sensory signaling through cranial nerves, relayed through brainstem nuclei (i.e. the nucleus of the solitary tract) and central measurement of homeostatic state. Despite these important caveats to the drive-reduction hypothesis, however, it is clear that activity in need-specific hypothalamic networks underlies motivation to eat or drink, and that this motivation follows from an aversive state.

Box 2:**Reinforcement learning**

In reinforcement learning, the goal is to learn a policy, $\pi(S_t)$, that maps states, S_t , into actions, $a_t = \pi(S_t)$. The optimal policy maximizes the sum of future (discounted) rewards, $V_{\pi}(s_t) = \sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t)$. (Note that a punishment can be coded as a negative reward.) The value of the current state under policy π is $V_{\pi}(S_t)$, and under the optimal policy this is the maximum reward that can be achieved over the relevant time horizon. State value can also be written as $V(s_t) = \max_{a_t \in A_{s_t}} \{Q(s_t, a_t)\}$. The variable $Q(s_t, a_t)$ is the action value, or the value of each action in the current state, and the set A_{s_t} includes all actions available in the current state. Action values can also be written as rewards and averages of next state values, $Q(s_t, a_t) = r(s_t, a_t) + \gamma \sum_{j \in S_{t+1}} p(j|s_t, a_t) V_{t+1}(j)$.

The policy, which is learned by RL, selects actions which maximize immediate and future rewards and minimize punishments. This incorporates the fact that making a decision that leads to a reward now may lead to punishments later. The discount factor, $\gamma < 1$, mediates the trade-off between current and future rewards and punishments. The optimal policy follows simply from having learned the values of actions available in each state, $Q(s_t, a_t)$, because once these are learned, the policy chooses the action with the highest value in each state, i.e. $a_t = \pi(s_t) = \operatorname{argmax}_{a_t \in A_{s_t}} Q(s_t, a_t)$. In other words, if I have accurately learned action values, $Q(s_t, a_t)$, I should pick the best action available in the current state. This also implies that the state defines all information necessary to select an action to maximize reward. Correspondingly, the state must contain all information necessary to act optimally. For biological agents, the state must contain internal variables (i.e. physiological variables like thirst or hunger that define current or future needs) and external variables (i.e. the environment - e.g. is food or water available), $s_t = \{S_{\text{internal}}, S_{\text{external}}\}$. Rewards, $r_t(s_t, a_t)$ also depend on the state, which reflects the fact that outcomes are rewarding only if they satisfy a current need, and the level of reward may reflect strongly the current internal state. When one is thirsty, a glass of water is much more rewarding than food. Because state contains information about both internal and external stimuli, it can map objects in the environment to the needs they satisfy. Thus, $V(s_t = \{\text{thirst}, \text{water available}\}) > V(s_t = \{\text{thirst}, \text{food available}\})$. However, reducing this to a scalar comparison hides the important fact that the critical element is knowing that a glass of water satisfies thirst, and a veggie burger satisfies hunger. This knowledge develops with experience and learning. Thus, the RL framework can be used to capture this information, but it should be done with the understanding of what is being learned.

The actor-critic model

The actor-critic framework is a specific implementation of RL theory. It provides an alternative interpretation, although not orthogonal to ours, of the dorsal and ventral circuits within the striatum [89, 90]. In neural implementations of the actor-critic framework, the critic represents state values in the VS and computes state value prediction errors in the dopamine neurons. The actor learns and implements the policy which resides in the DS [91]. State value and action value updates are driven by the

prediction errors computed by the critic. Aspects of this model are supported by data [91]. However, action value learning can take place in the absence of the VS [69], and thus the VS is not critical for the computation of prediction errors when learning action values. There are, however, several consistencies between the actor-critic model and the framework we are putting forward. Specifically we suggest that the VS, as part of the ventral circuitry, can represent state-values and the DS can represent action values. However, we suggest that: 1) a broader neural circuitry is involved in state-value representation; 2) the role of the hypothalamus is critical to computing state values, which are dependent on internal states and objects in the environment; and 3) action value learning, as it is often studied in the laboratory, does not play a prominent role in ethological behavior. It is possible that the dorsal circuitry, including the DS, may play a role in hierarchical reinforcement learning (see Outstanding Questions) [1, 92, 93].

Box 3:**The hippocampus and associated medial temporal lobe structures**

The hippocampus, from an anatomical point of view, is an interesting structure because it links the dorsal and ventral networks. In primates, parietal area 7a in the dorsal network (Fig. 1A) projects to the parahippocampal gyrus as well as retrosplenial cortex, both of which project to the hippocampus [94]. Retrosplenial cortex also projects to dorsal area 46. In addition, inferior temporal visual areas project to both entorhinal and perirhinal cortex which in turn project to the hippocampus. The hippocampus (including the subiculum in the designation hippocampus) then projects to vmPFC [95] and the ventral striatum [25], which brings it into the ventral, hypothalamic circuitry. The hippocampus also gets inputs, via the anterior thalamic nuclei, from the mamillary bodies in the hypothalamus [6]. So, the hippocampus receives spatial inputs from the dorsal system, as well as object-related visual inputs from the inferior temporal lobe, but its outputs interface with the ventral circuitry.

The hippocampus is often thought to be critical for episodic memory, although recent data suggests this simple interpretation may be incomplete [3, 96]. Given that the hippocampus and associated temporal lobe structures have a role in episodic memory, it has been suggested that these structures are important for learning state or action values in complex state spaces [97], in which one will visit each individual state only a few times. RL normally learns by repeatedly experiencing outcomes in each state. If the state space is large, rewards are delayed in time, and one only visits each state a few times, it will not be possible to learn effectively via the normal RL mechanisms. In this case, episodic memory can capture the details of important single events, and this memory can be used to make choices in the future.

Outstanding Questions

- Models of reinforcement learning often suggest that the dorsal striatum is important for learning action values. This is based on studies that use stereotyped laboratory actions that rarely lead to rewards in real life. Does the dorsal circuitry play a similar role in ethologically realistic conditions, or does the dorsal circuitry calculate actions on the fly to acquire goal-objects?
- Is learning in the dorsal system specialized for learning the hierarchical organization of action sequences (i.e. hierarchical reinforcement learning) that can be flexibly deployed to achieve goals?
- Neurophysiology studies that have recorded across connected nodes in the dorsal or ventral circuit often find similar responses. What distinguishes the contributions of the cortical, basal ganglia and thalamic nodes, all of which are defined by different microcircuit architectures? Is the synaptic or cellular plasticity that supports reinforcement learning present across all nodes, or is it predominantly restricted to specific nodes in each circuit (e.g. the striatum which is often assumed or the amygdala)?
- During reinforcement learning in the wild, is the neural plasticity that mediates learning equally distributed between the dorsal and ventral circuits? Or, is the plasticity primarily in the ventral circuit and therefore state-value learning?
- The basal ganglia are often thought to carry out action selection. However, neurophysiology studies usually find that the cortex discriminates actions before the basal ganglia. Does the basal ganglia do action selection? Similarly, does the basal ganglia select among competing motivations, or is this carried out in, for example, the lateral hypothalamus?
- Most learning studies focus on approach/avoid behaviors. However, subjects often learn need-specific values of goal objects. How are these need-specific values assigned to goal objects? For instance, how do we learn that a glass of water satisfies thirst and a veggie burger satisfies hunger?
- How are abstract goals, which can be far removed from immediate biological necessity, represented and learned?

Highlights

- We propose that the neural circuitry that underlies goal-directed behavior in primates is organized into two large-scale, cortical-subcortical neural systems, which we refer to as the dorsal and ventral circuits.
- The ventral circuit, due to substantial visual input from the medial temporal cortex and connectivity with the hypothalamus, learns about and identifies need-specific goal objects in the environment.
- The dorsal circuit, due to substantial parietal cortex input, represents spatial metrics relevant to actions. The dorsal circuit uses these representations to generate actions on the fly to obtain goal objects.
- Reinforcement learning provides a framework for modeling these separate functions of the dorsal and ventral circuits. The ventral circuit represents states (internal and external) and state values, and the dorsal circuit computes actions on the fly to obtain goal states, which have been identified by the ventral circuit.

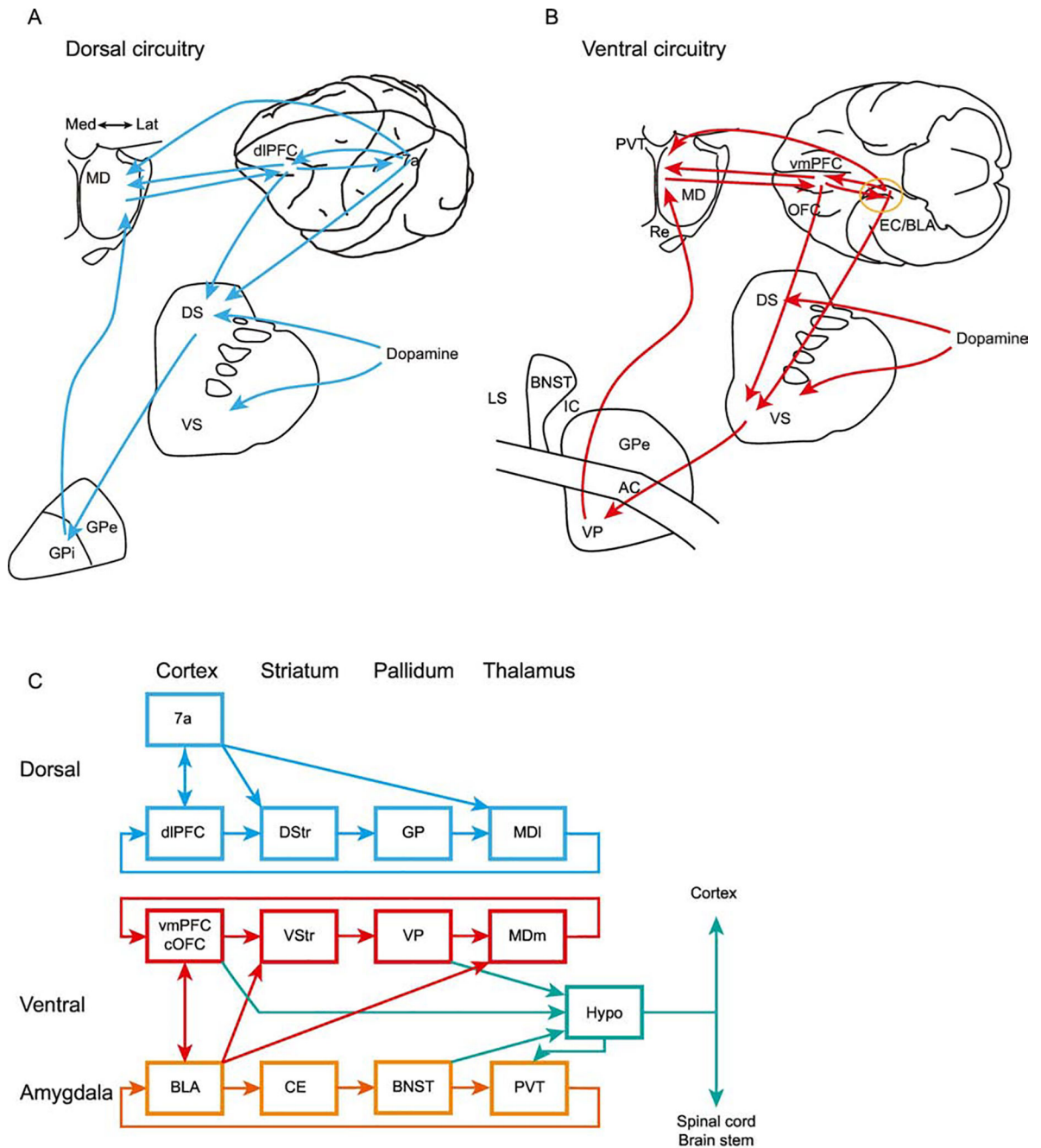
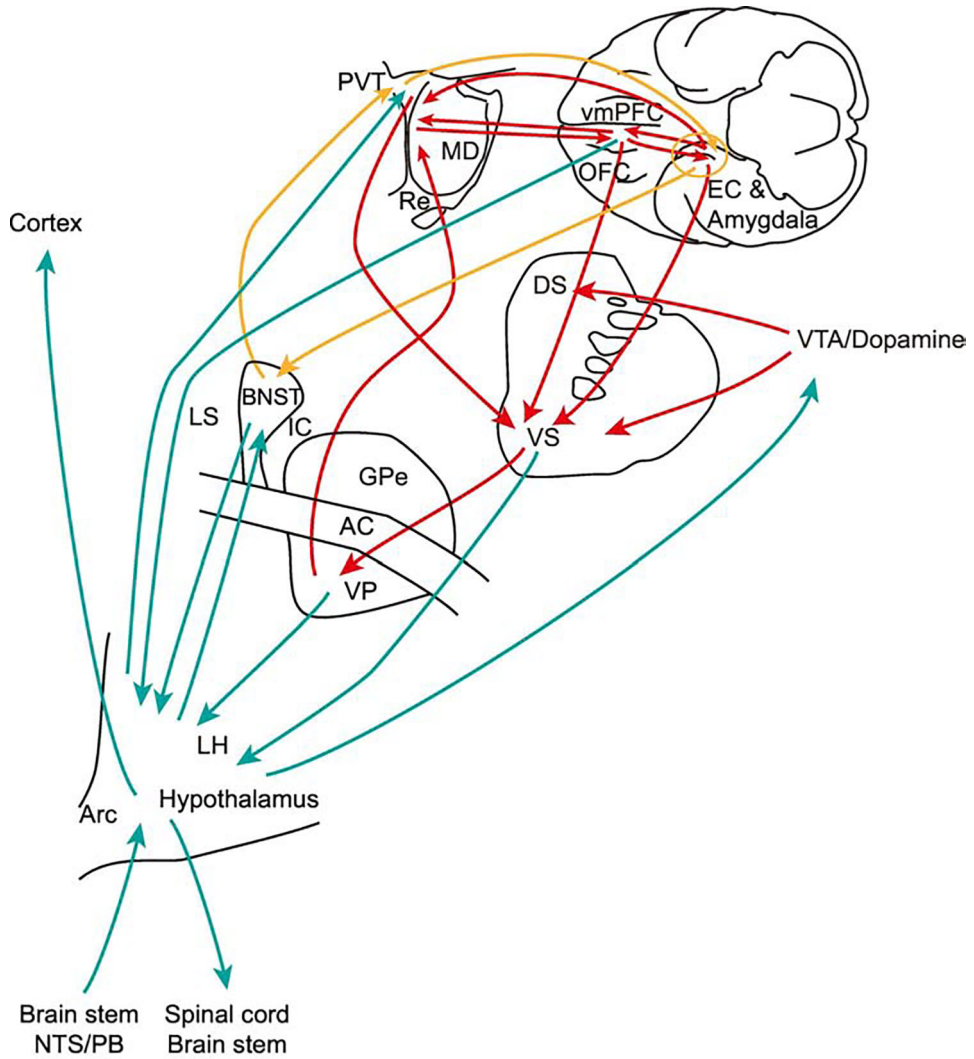


Fig. 1. Dorsal and ventral circuits in the primate brain for goal-directed behavior. A. Overview of the proposed dorsal circuit, displayed on schematics of the macaque brain (lateral view, top right), medial dorsal thalamus (top left), striatum (middle) and globus pallidus (bottom left). Dorsal circuit connecting inferior parietal cortex (area 7a) with dorsal lateral prefrontal cortex (dIPFC, area 46). The cortical circuits have overlapping projections in the dorsal striatum and lateral portion of the medial dorsal thalamus (MDI). The striatal circuit also connects through the globus pallidus internal segment (GPi). B. Overview of the proposal

ventral circuit (as in A, only with a ventral view of the brain in top right). The ventral circuit (red lines) connects the basolateral amygdala (BLA) and adjacent medial temporal lobe cortex including entorhinal (EC) and perirhinal (PrC) cortex to the ventral medial prefrontal cortex (vmPFC). The vmPFC and BLA/EC/PrC have convergent projections in the ventral striatum (VS) and medial MD thalamus. The VS connects through the ventral-pallidum (VP) back to the medial portion of the medial dorsal thalamic nucleus (MDm). The ventral circuit is generally consistent with the classical definition of the limbic system. However, considering the parallel organization of ventral and dorsal circuits may help understand their function. C. Schematic illustration of the dorsal, ventral and amygdala circuits. Many connections are not shown, for clarity. Note the substantial anatomical connectivity between the amygdala and the ventral circuitry. The ventral circuitry has substantial projections to the hypothalamus (Hypo), whereas the dorsal circuit has minimal connections to the hypothalamus. While the amygdala projects to the ventral circuit, it also has a parallel cortical-basal ganglia-thalamo-cortical architecture [30]. In this conception, the basolateral amygdala (BLA) is the (allo)cortical node. The BLA projects to the central nucleus (CeN) of the amygdala, which is part of the striatum. The CeN then projects to the bed nucleus of the stria terminalis (BNST; equivalent to the VP). The BNST projects to the paraventricular thalamus (PVT), which projects back to the amygdala, closing the loop [98]. There is also a hippocampal circuit, that largely parallels the amygdala circuit, which we have not shown for simplicity (see Box 3 for some comments on the hippocampus). This circuit projects through the lateral septum (LS) and medial septum, back through the reuniens (Re) nucleus of the thalamus.



Trends in Neurosciences

Fig. 2. Connectivity of ventral circuits with the hypothalamus. Colored lines match colors of each circuit from Fig. 1C. The vmPFC, BNST, VP and VS all send inputs to the hypothalamus. LH: lateral hypothalamus. Arc: arcuate nucleus of the hypothalamus. NTS: nucleus of the solitary tract; PB: parabrachial nucleus; see figure 1 for additional abbreviations. Note that we are illustrating connections from the surface of the medial temporal lobe. The basolateral amygdala, EC and PrC have similar connections.