

Prospective Optimization

This paper describes the idea that brains form cognitive strategies by prospective optimization, which is the planning of future actions to optimize rewards.

By TERRENCE J. SEJNOWSKI, *Fellow IEEE*, HOWARD POIZNER, GARY LYNCH,
SERGEI GEPSHTEIN, AND RALPH J. GREENSPAN

ABSTRACT | Human performance approaches that of an ideal observer and optimal actor in some perceptual and motor tasks. These optimal abilities depend on the capacity of the cerebral cortex to store an immense amount of information and to flexibly make rapid decisions. However, behavior only approaches these limits after a long period of learning while the cerebral cortex interacts with the basal ganglia, an ancient part of the vertebrate brain that is responsible for learning sequences of actions directed toward achieving goals. Progress has been made in understanding the algorithms used by the brain during reinforcement learning, which is an online approximation of dynamic programming. Humans also make plans that depend on past experience by simulating different scenarios, which is called *prospective optimization*. The same brain structures in the cortex and basal ganglia that are active online during optimal behavior are also active offline during prospective optimization. The emergence of general principles and algorithms for goal-directed behavior has consequences for the development of autonomous devices in engineering applications.

KEYWORDS | Basal ganglia; cerebral cortex; classical conditioning; dynamic programming; hippocampus; ideal observer;

limbic system; optimization; reinforcement learning; temporal-difference learning

I. INTRODUCTION

Bellman's approach to optimizing a sequence of actions to reach a goal is based on known state transitions and payoffs [1]. Dynamic programming is an effective strategy for moderately sized problems, but the complexity increases rapidly with the size of the problem, leading to the "curse of dimensionality." For an animal that is exploring an unknown environment with many choices for actions and limited knowledge this is not a feasible strategy, although the goal of optimizing future rewards is the same.

Reinforcement learning is an online approach to dynamic programming that proceeds incrementally to build a value function that can be used to choose optimal actions. In classical conditioning, found in many species, associations are learned between sensory stimuli and rewards depending on the order and timing of the pairings, as described by the Rescorla–Wagner model of classical conditioning [2]. The temporal-differences algorithm in reinforcement learning is closely related to the Rescorla–Wagner model [3], [4], and approximates dynamic programming [5]. This approach constructs a consistent value function for states and actions based on feedback from the environment. Classical conditioning, which might seem like a weak way to learn about the world, could thus lead to near-optimal strategies for finding food, shelter, and mates over many trials [6].

An impressive demonstration that reinforcement learning can solve difficult problems is TD-Gammon, a program that started as a beginner and improved by playing itself, eventually achieving world champion level in backgammon [7]. Solely on the basis of the reward at the end of each game, TD-Gammon discovered new strategies that had eluded experts. This illustrates the ability of reinforcement learning to solve the temporal credit assignment problem and learn complex strategies that

Manuscript received February 4, 2014; revised March 20, 2014; accepted March 24, 2014. Date of publication April 22, 2014; date of current version April 28, 2014. This work was supported by the U.S. Office of Naval Research (ONR) under the Multidisciplinary University Research Initiative (MURI) Grant N000141010072; by the National Institutes of Health (NIH) under Grants NS036449 and EY018613; and by the National Science Foundation (NSF) under Grants #114670811, SBE-0542013, and ENG-1137279. The work of T. J. Sejnowski was also supported by the Howard Hughes Medical Institute.

T. J. Sejnowski is with the Howard Hughes Medical Institute, Salk Institute for Biological Sciences, La Jolla, CA 92037 USA and the Division of Biological Studies, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: terry@salk.edu).

H. Poizner is with the Institute for Neural Computation, University of California at San Diego, La Jolla, CA 92093-0523 USA (e-mail: hpoizner@ucsd.edu).

G. Lynch is with the Department of Psychiatry and Human Behavior, University of California at Irvine, Irvine, CA 92697-4292 USA (e-mail: glynych@uci.edu).

S. Gepshtein is with the Systems Neurobiology Laboratories, Salk Institute for Biological Studies, University of California at San Diego, La Jolla, CA 92037 USA (e-mail: sergei@salk.edu).

R. J. Greenspan is with the Kavli Institute for Brain and Mind, University of California at San Diego, La Jolla, CA 92093-0126 USA (e-mail: rgreenspan@ucsd.edu).

Digital Object Identifier: 10.1109/JPROC.2014.2314297

0018-9219 © 2014 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

lead to winning ways. Reinforcement learning has also been used to learn complex control laws. For example, flying a helicopter is much more difficult than flying an airplane, but a control system was trained with reinforcement learning to perform helicopter aerobatics [8]. However, despite these notable successes, reinforcement learning does not always converge to good solutions for control problems.

Nature has integrated reinforcement learning with other brain systems to handle the complexity of the world and the limited number of choices that an animal can make in its lifetime: Early brain development creates the basic patterns of wiring in the brain and experience during life modifies these connections [9]; the declarative memory system involving the hippocampus and structures in the medial temporal lobes of the cortex allows memories of specific events and objects to be accessed to guide behavior [10]; the complexity of a problem is reduced by limiting the number of sensory stimuli that are attended at any given time [11]; finally, cognitive systems evolved to plan future strategies.

In this review, we first consider conditions under which performances on sensory and motor tasks achieve near optimality. Successful behavior depends on the ability to link expected outcomes of action with relevant models of the world, to derive expectations of reward. In the next part, this will be illustrated by the problem of learning where to look, which depends on gathering information from the world over time to achieve an optimal search strategy. These first two sections are focused on behavior and how performance is shaped by experience. In the third section, the areas of the brain involved in reinforcement learning are introduced, which are responsible for organizing sequences of actions to reach goals. The anatomical organization of these areas, and in particular the loops between the cerebral cortex and the basal ganglia, reveal levels of control that make us more flexible and adaptable. Finally, we examine how brains form cognitive strategies by prospective optimization—planning future actions to optimize rewards. These more advanced aspects of reinforcement learning have the potential to greatly enhance the performance of autonomous control systems.

II. IDEAL OBSERVERS AND PERFORMERS

Human performance on most tasks is rarely optimal and a great deal of learning is required to achieve good performance. However, in some tasks, ranging from sensory perception to decision making, humans and other species can perform at or near the limit that is determined by an “ideal observer” and “ideal performer” that has access to all the essential information and performs perfect inference on that information [12], [13]. For example, the detection of photons at the photoreceptor is at the noise

limit set by signal detection theory [14], [15] and imperfections of neuronal communication [14], [16]–[20]. Perceptual grouping and contour detection in humans is also near the performance of an ideal observer [21]. An example of a search task is given in Section III where humans perform near the ideal limit after learning. This suggests either that nature is performing Bayesian inference on joint probability distributions, which requires immense memory resources, or some approximate scheme that approaches optimal performance [22], [23].

Perceptual experiments allow us to probe the mechanisms that may be responsible for achieving near-optimal performance. Learning experiments provide evidence for how optimal performance is acquired after extended experience with the environment. Brain recordings during these tasks have indicated that interactions between the cerebral cortex and the basal ganglia are involved in learning new skills and achieving near-optimal performances, which will be discussed below.

In addition to observing and acting, humans also excel in planning future actions. Prospectively imagining the future activates many of the same brain areas that are engaged in remembering the past. These include regions of the medial prefrontal cortex, hippocampus, and posterior regions of the parietal cortices. Thus, memory of past events can be used to generate possible future events. We examine evidence for this ability from rodent to human studies, including results from patients with Parkinson’s disease, and the key role of the basal ganglia and the dopaminergic system in prospection. Although we have a good theoretical understanding of reinforcement learning and the neural circuits underlying it in the basal ganglia, the complexity of prospective optimization requires a new conceptual framework that includes interactions of the basal ganglia and hippocampus with the prefrontal cortex.

Many models are concerned with uncertainty (the degree of precision) of sensorimotor computations. For example, in studies of visual depth perception, information was shown to combine according to its relative precision [24], [25], and likewise for vision and touch [26] or vision and hearing [27]. In these models, information from different cues is weighted according to the variance of corresponding sensory estimates. More precise cues, such as cues with smaller variance, were shown to contribute to perception more than the less precise cues so that perceptual contributions of the cues changed accordingly [27], [28].

But not all predictive theories of sensory processes are statistical: A theory of human spatio-temporal sensitivity [29] recognizes the fact that selectivity of receptive fields in sensory neurons limits the information they communicate [29], [30], and offers prescriptions for how distributions of receptive fields can be optimally matched to the environment; some theories of perception of visual shape assume that vision favors simple organization [31]–[34].

In the realm of motor behavior, the limiting uncertainty concerns precision of motor acts, e.g., pointing, reaching, and grasping hand movements. For example, in a study of rapid reaching movements toward small visual stimuli [35], touching of overlapping disks (“target” and “penalty”) incurred, respectively, monetary rewards and penalties. Human subjects tried to maximize payoffs by aiming to the side of the target region away from the penalty. Taking into account the motor uncertainty measured separately in every subject, a normative model (“ideal planner”) predicted that the greater the uncertainty, the larger the expected shift of aim point away from target center. Confirmation of this expectation supported the view that human neural systems take into account task-relevant motor uncertainties in planning action.

Subjects showed considerable flexibility in this task. When the visual feedback was manipulated to create an impression that the scatter of movement end points was larger than it really was, subjects changed their aim points in agreement with quantitative predictions of the ideal planner [36]. Similarly, when movements were directed to multiple stimuli in rapid succession, in different parts of the visual field, such that shapes of 2-D spatial distributions of end points (“shapes of motor uncertainty”) were different for different stimuli, subjects were able to adjust their aim points accordingly, as also predicted by the ideal planner [29]. Thus, human neural systems are able rapidly to evoke representations of uncertainty that match the immediate task.

How can such optimization of action be achieved? Action planning has a retrospective aspect that builds upon previous interactions of the organism and its environment, i.e., “sensory adaptation” and “motor adaptation.” But learning about present sensory and motor uncertainties could also have beneficial value in the future. Hence, it is important to model prospective aspects of action planning, i.e., “prospective optimization.” Retrospective and prospective aspects of action planning are tightly intertwined, rather than being separate processes. As agents carry out extended actions, they are learning the context of action and apply this knowledge toward making future decisions.

The aforementioned ideal-planner models implement some features of prospective optimization: computing gains expected from different actions and selecting actions whose expected gains are largest. For example, in the task with two overlapping disks of reward and penalty, the model computes the gains for every point in a large area that includes the disks, thus yielding the “gain landscape” for this task. The spatial location where the landscape reaches its peak constitutes the prediction of optimal aim point.

Yet these models are still incomplete. They fail to capture two essential features of biological prospective optimization: the natural environment is highly dynamic and computational capacities of neural systems are limited: their computational “reach” (horizon) into the future is bounded. These two features of realistic action

planning interact. In the dynamic environment, new properties of the environment are continuously revealed to the organism as they enter the scope of computation. Accordingly, the expected gains of action must be continuously reevaluated, and this capability needs to be incorporated into the normative models.

III. LEARNING WHERE TO LOOK

You are approaching a road and look to your left before stepping out. This strategy is effective in North America, but can be fatal in the United Kingdom. As you walk along a road you know where to look for streets signs and street addresses. Knowing where to look when searching for information in an environment is highly context sensitive and learning where to look in a dynamic environment is a good example of prospective optimization.

Learning where to look involves the evaluation of sensory information, a form of “bottom-up” processing, integrated with attentional processes driven by “top-down” expectation. These two processes are intertwined in the brain and are difficult to disentangle, but recently a novel search task was developed to tease them apart [37]. Participants were seated in front of a blank screen and told that their task was to explore the screen to find a hidden target location that would sound a reward tone when fixated. The hidden target position varied from trial to trial and was drawn from a Gaussian distribution not known to the participant but held constant throughout a session [see Fig. 1(a)].

At the start of a session, participants had no prior knowledge to inform their search. Once a fixation was rewarded, participants could use that feedback to assist on the next trial. As the session proceeded, participants could improve their success rate by developing an expectation for the distribution of hidden targets and using it to guide future search [Fig. 1(a)]. After remarkably few trials, participants gathered enough information about the target distribution to efficiently direct gaze, as illustrated by one participant’s data in Fig. 1(a) and (b). After approximately a dozen trials, fixations narrowed to the region with high target probability. A characterization of this effect for all participants is shown in Fig. 1(c). The search spread was initially broad and narrowed as the session progressed, as shown in Fig. 1(d).

An ideal observer was derived for this task assuming that fixations are independent of one another and that the target distribution is known. The dashed lines in Fig. 1(a)–(c) mark ideal-observer performance. Optimal search performance requires a distribution of planned fixation “guesses” that is approximately $\sqrt{2}$ broader than the target distribution itself [38], [39]. As seen in Fig. 1(b) and (c), the performance of participants hovered around this optimal search distribution after about a dozen trials. In Fig. 1(a), the mean for the human data from trials 31–60 is higher than the theory suggests, but the theory presumes

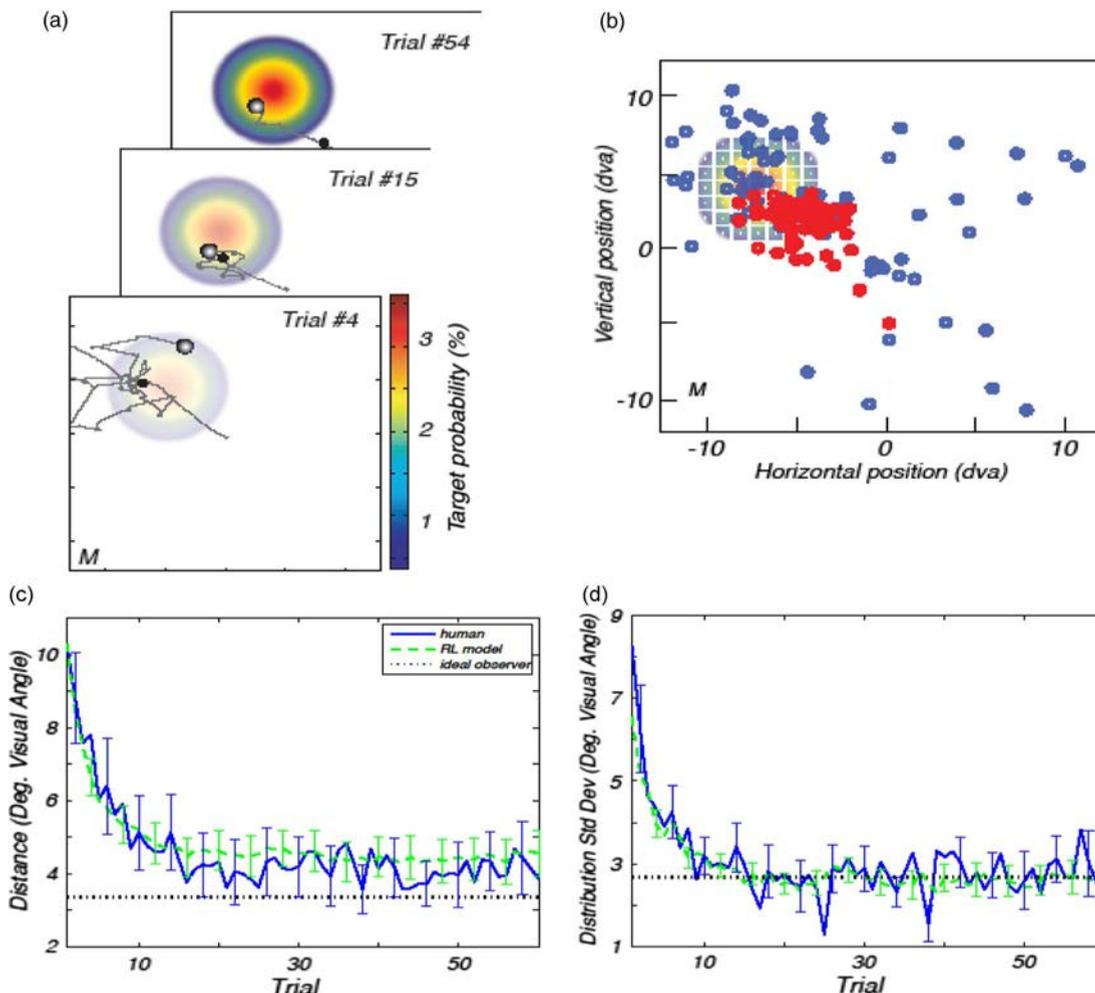


Fig. 1. Hidden target task. (a) Blank screen is superimposed with the hidden target distribution that is learned over the session as well as sample eye traces from three trials for a participant. The first fixation of each trial is marked with a black circle. The final and rewarded fixation is marked by a shaded gray-scale circle. (b) The region of the screen sampled with fixation shrinks from the entire screen on early trials (blue circles; 87 fixations over the first five trials) to a region that approximates the size and position of the Gaussian-integer-distributed target locations on later trials (red circles; 85 fixations from trials 32–39). (c) Learning curves. The distance between the mean of the fixation cluster for each trial to the target centroid, averaged across participants, is shown in blue and green and indicates the result of 200 simulations of the reinforcement-learning model for each participant’s parameters. The standard error of the mean is given for both. The ideal-observer prediction is indicated by the black dotted line. (d) The standard deviation of the eye position distributions or “search spread” is shown for the average of all participants (blue) and the reinforcement-learning model (green) with standard error of the mean. The dashed line is the ideal-observer theoretical optimum in each case, assuming perfect knowledge of the target distribution. (Adapted from [37].)

stationarity of the target distribution. Individuals must be responsive to nonstationarities in natural environments and this responsivity yields an increase in uncertainty [40] consistent with observed human performance.

In addition to the ideal-observer theory, the task was also modeled using a temporal-difference algorithm in reinforcement learning [41], which reduces the error of predicted future rewards and is motivated by animal learning and behavioral experiments [42]. This model constructs a value function mapping locations in space to expected reward. The value function is updated after each fixation based on whether the target is found, and is used

for selecting saccade destinations that are likely to be rewarded. Two additional assumptions were made: First, each time a saccade is made to a location, the feedback obtained generalized to nearby spatial locations; second, humans tend to make more short saccades than long saccades, which was incorporated in the value function as a proximity bias. Because the choice of the next fixation became dependent on the current fixation, an ideal observer would plan sequences of fixations instead of choosing a set of independent fixations.

The mean performance of the model closely tracked mean human performance [Fig. 1(c)]. The model also

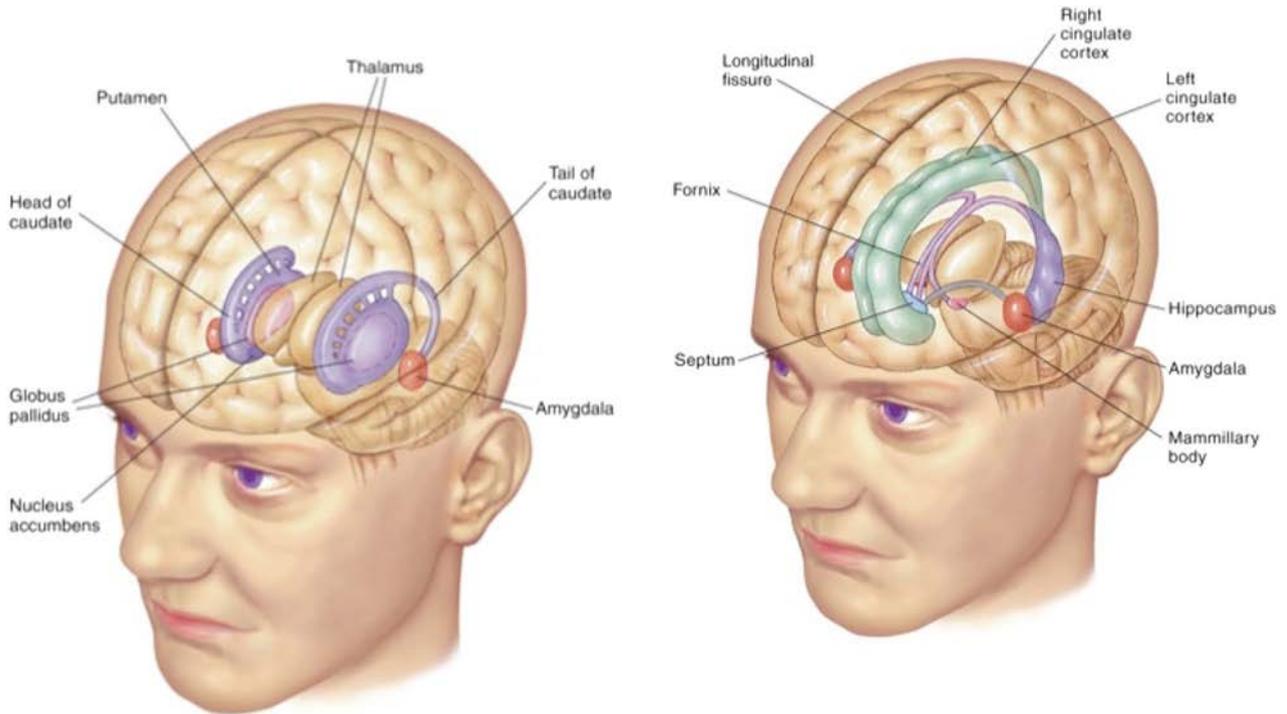


Fig. 2. Brain structures comprising the (left) basal ganglia and (right) limbic system. The cerebral cortex shown on the outside of the brain projects to the caudate and putamen of the dorsal (top) striatum, and nucleus accumbens of the ventral (bottom) striatum of the basal ganglia. The output of the basal ganglia from the globus pallidus projects to the thalamus, which then project back to the cortex, forming a loop (see Fig. 3). The limbic system, which means “ring” and circles the thalamus, regulates emotion, behavior, motivation, long-term memory, and olfaction. The limbic system includes the cingulate cortex on the inside, or medial wall of the cortex, the hippocampus and the amygdala. (Courtesy of Paul Wissmann.)

predicted an asymptotic search spread that increased with the target spread, consistent with aggregate performance [Fig. 1(d)]. Similar to the human performance observed in Fig. 1(c), the reinforcement-learning model approaches, but does not reach, the theoretical asymptote. Like the human participants, reinforcement-learning model was responsive to nonstationarity in the distribution, whereas the ideal-observer theory assumes that the distribution is static.

The success of temporal-difference learning and ideal-observer theory raises the question of how prospective optimization is actually implemented in nervous systems. The neuroanatomy of the brain is complex, and we only have a crude understanding of the function of most brain areas. Progress has been made recently in understanding the role of dopamine, a neuromodulator that is associated with reward systems, in guiding actions through its influence on the cerebral cortex and the basal ganglia.

IV. DOPAMINE NEURONS AND REWARD-PREDICTION ERROR

The basic components of the basal ganglia, shown in Fig. 2(a), can be identified in the lamprey, a representative

of a group that emerged near the beginning of vertebrate evolution, suggesting that this system is phylogenetically ancient [43]. The basal ganglia receive inputs from most of the cortical mantle and in turn project back to the cortex, through the thalamus, forming long parallel loops. The dorsal basal ganglia are regions that organize voluntary motor control, in selecting actions and learning sequences of actions. The ventral basal ganglia, which receive projections from the frontal cortex, have been implicated in higher cognitive functions and emotional control. The dorsal and ventral basal ganglia are heavily innervated by inputs from dopamine neurons from the substantia nigra pars compacta or ventral tegmental area, which are involved in rewards and reinforcement learning. Stimulation of dopamine pathways mediate the rewarding effects of intracranial self-stimulation [44].

The basal ganglia compute the predicted reward for the current state of the world represented in the cortex and compare it with the actual reward that is received; the difference between the expected and received reward is signaled by transient changes in the firing rates of dopamine neurons, which is then used to update the prediction through changes in the strengths of cortico-striatal synapses [45], [46]. The same dopamine signal can

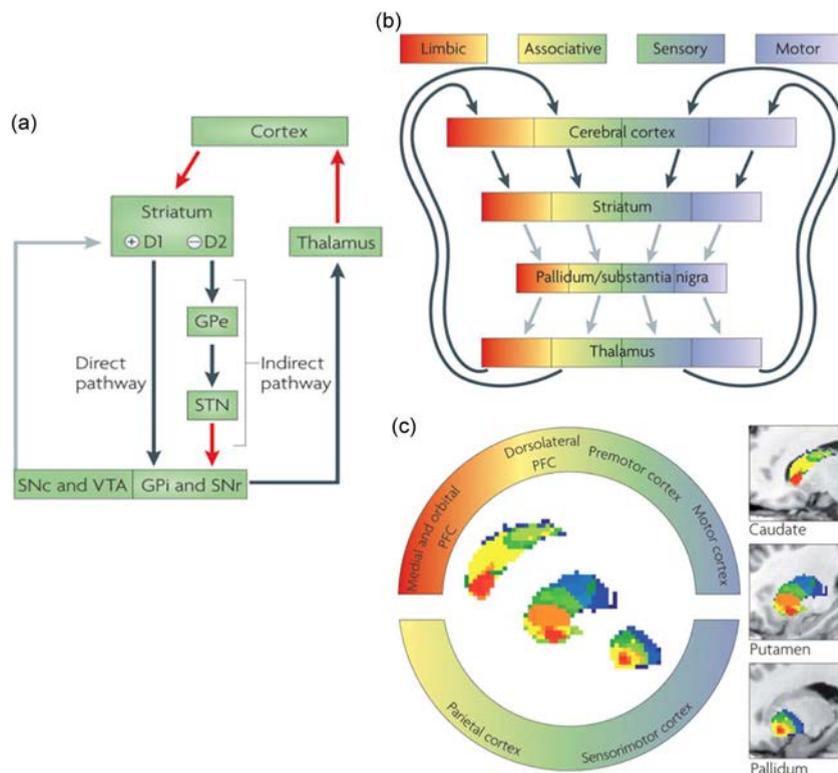


Fig. 3. Schematic model of cortico-striatal loops. (Left) Model of the basal ganglia showing the direct pathway—which involves direct striatonigral inhibitory connections (dark green arrows) that promote behavior—and the indirect pathway—which involves relays in the external globus pallidus (GPe) and subthalamic nucleus (STN), with the only excitatory projection in the basal ganglia (red arrow), and suppresses behavior. The balance between these two projections is thought to be regulated by afferent dopaminergic signals from the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA). (Top right) The connections between the cerebral cortex and the basal ganglia can be viewed as a series of parallel-projecting, largely segregated loops or channels conveying limbic (red), associative (yellow-green) and sensorimotor (blue-white) information. Functional territories represented at the level of cerebral cortex are maintained throughout the basal ganglia nuclei and thalamic relays. Black arrows indicate excitatory glutamatergic projections, gray arrows indicate GABA-ergic projections. (Bottom right) The spatially segregated “rostral caudal gradient” of human prefrontal cortical connectivity in the caudate, putamen, and pallidum. The color-coded ring denotes limbic (red), associative (yellow-green) and sensorimotor regions of the cerebral cortex in the sagittal plane. PFC: prefrontal cortex. (Adapted from [56].)

be used to make decisions: Each possible action is considered in turn and the one that elicits the highest level of dopamine is chosen. The discovery that transient dopamine signals indicate reward-prediction error has given rise to new insights into how decisions and plans are made. Here we will explore how the interactions between the cerebral cortex and the basal ganglia contribute to efficient planning of future actions [6].

Findings over many years and in many species have led to the view that the architecture of the basal ganglia contains multiple, closed feedback loops, linking striatal zones with cortical regions, in which the dorsolateral stream executes sensorimotor functions, while the ventromedial stream is more closely related to motivations and emotions (see Fig. 3) and interact with the limbic system [Fig. 2(b)]. Dopaminergic modulation of activity within the ventral striatum also has a potent influence on diffuse ascending systems that regulate emotion through the hypothalamus and higher limbic structures. The habenula,

a phylogenetically ancient structure that receives inputs from many limbic structures, inhibits dopamine neurons in the substantia nigra pars compacta and also influences neuromodulatory nuclei of the brainstem, and is a source of negative reinforcement signals in dopamine neurons [47].

Dopamine cells in the primate substantia nigra pars compacta predict decisions for future actions [48], which is related to reward prediction error. Dopamine population activity is modulated according to the future actions of the monkey rather than to the reward probability itself. The ability to choose future actions under conditions of dopamine depletion has been studied in patients with Parkinson’s disease (PD) who, when they are not taking their dopaminergic therapy, showed deficits in learning to choose optimal actions, most acutely at the point in time when the reward probabilities changed [49]. Thus, the nigral-striatal dopaminergic system seems to be critical for optimizing our decisions for future action in stochastic environments.

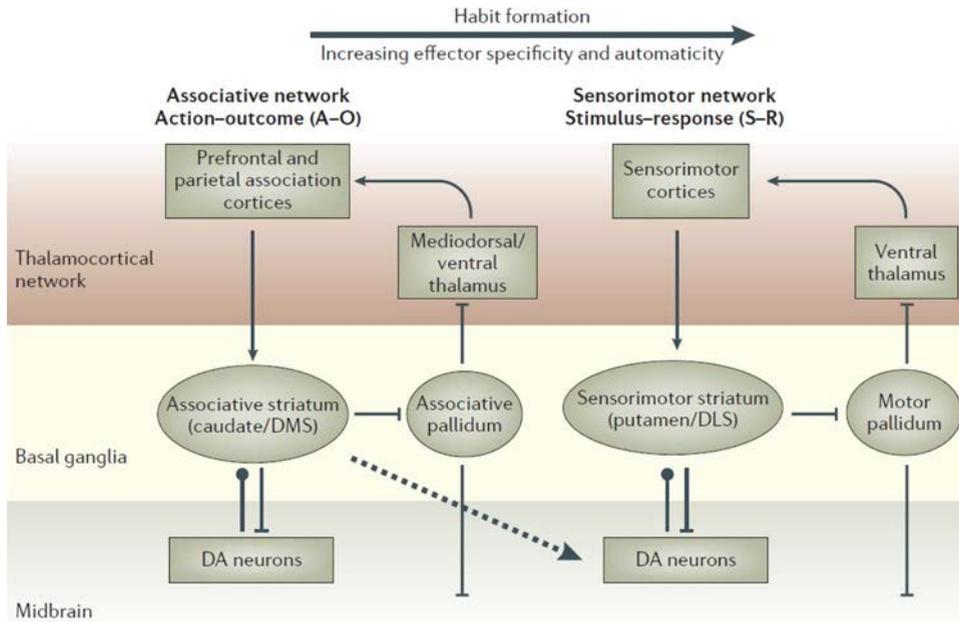


Fig. 4. Organization of cortico-basal ganglia networks. Schematic illustration showing cortico-basal ganglia networks in relation to serial adaptation. A shift from the associative to the sensorimotor cortico-basal ganglia network is observed during habit formation. DA: dopamine; DLS: dorsolateral striatum; DMS: dorsomedial striatum. (Adapted from [57].)

Optimization for future actions must not only take into account the magnitude of the reward, but also the delay until the reward becomes available [50]. In animals and humans, the subjective value of a reward decays hyperbolically, the longer the delay [51]. In rats, the responses of dopamine neurons show a similar hyperbolic decay function to reward delay [52]. Uniquely, humans discount smaller reward amounts more steeply than larger amounts [51]. Discounting itself is affected by basal dopamine levels, as indicated by studies on patients with PD. When some PD patients are given dopaminergic therapy, they develop pathological gambling or other impulse control disorders [53], presumably due to “overdosing” dopamine levels in a relatively intact ventral striatum [54]. PD patients with impulse control disorders strongly prefer immediate, small rewards over delayed larger rewards [55], consistent with a high discount.

V. ELABORATIONS OF BASIC CIRCUITS AND PROSPECTIVE CAPABILITIES

The basal ganglia and the cerebral cortex have many different regions corresponding to different sensory systems, different aspects of planning and motor control, and their corresponding working memories, which are used to maintain relevant information while performing a task. In this section, we will explore the functional roles that each of these regions may have in learning to control behavior. The anatomical terminology is, unfortunately,

arcane, but conceptually the overall goal of these regions is clear: keep track of the state of the world, predict the outcomes of possible actions, and decide which actions to take.

The putamen of the striatum receives inputs from sensory and motor regions of the cortex and is involved with habit formation [blue loop in Fig. 3(b)]. This loop maps sensory states to responses (S-R), reinforced by rewards. In reinforcement-learning theory, this is called model-free learning, in which the basal ganglia serves as a lookup table that associates brain states values and actions [57]. This strategy is closely tied to the effector (such as a hand) and can make decisions rapidly. However, a habit is an inflexible strategy that does not allow for contingencies. In particular, reward contingencies can change more rapidly than can be accommodated by habit formation.

The loop through the caudate of the striatum, which receives inputs from the dorsolateral prefrontal cortex and other associative areas of the cortex [yellow-green in Fig. 3(b)], is linked to outcomes rather than responses; that is, learning is directed toward associating a particular action to a particular reward (A-O), as shown in Fig. 4. In reinforcement-learning theory, the A-O system is a model-based approach, in which the action is linked to outcomes by constructing a model of the environment. The dorsolateral prefrontal cortex also supports working memory, which is a much faster memory system that can rapidly adapt to task contingencies. There is a hierarchy in learning in which a novel task initially under the control of

the more flexible associational loop is transferred to the sensorimotor loop (Fig. 4).

A third cortico–striatal loop through the ventral striatum receives inputs from the orbitofrontal and ventromedial regions of the cortex [red loop of Fig. 3(b)]. The most striking change to the basic dopamine–striatal networks across the vertebrates is the increase of connections that link ventral striatum to prefrontal cortex, as a dense projection from the ventral pallidum to the mediodorsal thalamus, which then innervates much of the prefrontal cortex. Reward prediction becomes elaborated in this loop to deal with complex environments that change in response to goal-directed behavior, an essential feature of fully realized prospective optimization. The habenula, which carries negative reinforcement signals, inhibits dopamine neurons in the ventral tegmental area, which in turn project to the ventral striatum.

The orbitofrontal and related medial prefrontal areas projecting to the ventral striatum are critical for the evaluation of delayed rewards [58] and orbitofrontal disturbances in rats and primates interfere with evaluation of the relationship between delay and reward value, and produce an exaggerated preference for immediate versus delayed rewards. Thus, the prefrontal areas enhance fundamental, dopamine-related ingredients of temporal discounting. Two well-established prefrontal cortical operations—working memory and sequencing of behaviors—are also likely to extend the reach of prospective computations. A high capacity working memory system would allow an increase in the number of cue elements that can be used in calculating the relative advantages of immediate versus delayed rewards, as well as presumably extending the time frame over which such calculations can be made. Enhanced ability to plan action sequences could, when coupled with expanded working memory, allow a brain to compare multiple, potential trajectories to a distant goal [6], [59].

The essential components of these three loops appear to be present in all mammals [39]. However, while the elements of the loop are conserved, the balance between them changes drastically with increases in brain size. The relative size of the cortex in primate species also expands more rapidly than the basal ganglia; moreover, some prefrontal areas expand disproportionately with increases in cortical size [60].

VI. A CONCEPTUAL FRAMEWORK FOR PROSPECTIVE OPTIMIZATION

There is a cost to pay for evaluating possible actions as they increase in number and complexity. It is relatively easy to make a decision when there are only two possible choices, but when there are many choices to make, time and computational resources become limited. One approach is to add the cost of having a complex policy into the value of future reward. This leads to a modification of temporal-

difference learning in which the decisions are based on an experience-modulated version of the behavior policy [61]. However, this approach becomes problematic when imagined scenarios are included in the set of possible choices, which could lead to an unending sequence of comparisons. Sutton has nonetheless shown that including some imagined scenarios can improve performance of reinforcement learning [62].

A system capable of *a priori* comparisons of multiple decisions and of using past credit assignments to choose between them, however, would not necessarily be capable of switching strategies once the action sequence was underway—a key element of prospective optimization. Damage to prefrontal areas outside the orbital cortex disrupts the ability of rodents and primates to switch between classes of discriminative cues (e.g., tactile to visual) while performing complex tasks [63]–[66]. The attentional “set shifting” analyzed in these studies is a logical early step in changing strategies when unexpected consequences arise during the execution of a planned trajectory. This leads naturally to including areas of the prefrontal cortex in the loop with the basal ganglia evaluating actions online. Set shifting engages multiple regions of the prefrontal cortex, which when damaged lead to perseveration. Thus, the interactions between regions of the prefrontal cortex are central for prospective optimization, but are not sufficiently well studied in humans to develop a model.

Recent work on the hippocampus and allied temporal lobe structures suggests that these areas play critical roles in episodic memory, very likely including the replay of already learned material. There are now reasons to suspect that the hippocampus has a similar role in prospectively organizing learned material pertinent to actions yet to be committed. A striking example of this was obtained with multielectrode recordings from the hippocampus of rats dealing with a multichoice problem; the spatio–temporal firing patterns at choice points provided a representation of first one and then the other of the potential response trajectories [67]. Brain imaging studies indicate that the anterior hippocampus is activated to different degrees while subjects are asked to imagine future events of varying likelihoods of occurrence [68]. Another recent fMRI study showed that medial prefrontal activation predictive of the valuation assigned to future rewards was associated with enhanced coupling of prefrontal–hippocampal activity, thus providing evidence that prefrontal cortex uses information from hippocampus for temporal discounting [69].

The prefrontal cortex also is bidirectionally coupled with the basal ganglia in ways that change with movement and dopaminergic input [70]. Moreover, during active navigation and decision making in rats, the striatum and the hippocampus likewise are functionally coupled [71]. Cross-frequency-band coupling changes dynamically both within and across striatum and hippocampus, particularly

during decision-making epochs, when simultaneous activation of synchronized striatal and hippocampal memory circuits occurs [71]. The ventral striatum is strongly implicated in delay discounting [72], [73], and, importantly, neuron firing in the ventral striatum is directly modulated by the hippocampus [74], [75]. Episodic future thinking reduces temporal delay discounting, in part by modulating networks involving the hippocampus [76]. Indeed, patients with hippocampal amnesia lose the ability to imagine novel experiences [77], reducing their ability to make optimal decisions [78]. Healthy humans have particularly long time lines for planning into the future and waiting long periods in order to achieve goals.

Our ability to imagine the future may put a brake on temporal discounting and impulsive behavior, promoting cooperation and constraint, operations that are particularly advantageous in highly interdependent human societies [79]. Indeed, the emergence of prospective thinking may have coincided with the emergence of a rapid expansion of behavioral repertoires that culminated in *homo ergaster/erectus* and *homo sapiens* [80]. Less complex brains can easily accomplish retrospective optimization, but the longer into the future that actions must be planned, the greater the brain complexity may be required. The ability for prospective optimization may have thus been an important driver in the evolution of increasing brain complexity.

These observations suggest modifications to working model that incorporate prospecting. Access to the unique anatomical machinery and prospective operations of hippocampus would allow prefrontal cortex to substitute imagined episodes for simpler inputs and thereby incorporate a “likely energy to be spent” component into temporal discount calculations. The apparent ability of hippocampus to serially generate representations of response sequences yet to be performed would permit the sequencing functions of prefrontal cortex to deal with much more complex possibilities (e.g., trajectory A followed by trajectory B) than would otherwise be possible. Forward and backward replays of place cells have been observed in the rat hippocampus [81], [82].

We expect that prospective operations will be found at multiple levels of the cortical mantle, with types of information and depth of extension into the future depending on local anatomy coupled with extrinsic connections. Final decisions, we further argue, depend on a prefrontal cortex that utilizes its close relationship with phylogenetically older systems to execute the time demanding calculations needed for prospective optimization. The extraordinary expansion of these prefrontal areas in humans would therefore allow for the remarkable efficiency of humans dealing with uncertain outcomes.

In humans it is possible explicitly to ask subjects about the past and the future. When asked to imagine future scenarios, the regions of the cerebral cortex that are activated are the same ones that are activated when asked

to remember past episodes [83], [84]. These areas include the medial and lateral temporal lobes, the lateral parietal cortex, the medial prefrontal cortex and the precuneus/retrosplenial cortex, as well as the hippocampus and the parahippocampal gyrus, which collectively form a core set of brain regions that are engaged in both verbally guided remembering and planning.

VII. CONCLUSION

Imagining the future and modifying behavior accordingly is one of the most adaptive capabilities of neural systems, especially the ability to use memory of past events to generate expectations of future events. Prospective optimization has become highly elaborated as the cortex and basal ganglia evolved to support increasingly longer time horizons and more complex behaviors. We have presented a conceptual framework for how prospective optimization may be integrated into the existing dopamine framework for reward prediction in the basal ganglia. Evolution has integrated all of these brain systems in ways that we are just beginning to appreciate.

As we continue to dissect the complexity of the circuits in the cerebral cortex and basal ganglia, as well as other parts of the brain such as the cerebellum that bring additional capabilities, it should be possible to elaborate on current reinforcement-learning systems to improve their performance on many practical problems. A new field called “dynamic cognitive systems” combines reinforcement learning with statistical signal processing and information theory to improve the performance of radios, radar, control, and power grids [85], [86]. In particular, these systems might benefit from including prospective optimization as an integral part of their architecture. As these engineered systems are developed and deployed based on these principles, they may in turn give us further insights into the cognitive aspects of brain function.

GLOSSARY

Navigational terms: The cardinal directions in the brain are dorsal (front), ventral (back), medial (closer to the midline), lateral (farther from the midline), rostral (anterior), and caudal (posterior).

Basal ganglia: Group of nuclei in the forebrain associated with voluntary motor control, procedural learning relating to routine behaviors or “habits,” eye movements, cognitive, and emotional functions (see Fig. 2). Basal ganglia are involved in action selection; that is, which of several possible behaviors to execute at a given time

Cingulate cortex: Cortical area forming a belt on the medial walls of the hemisphere; part of the limbic system and involved with linking behavioral outcomes to motivation and emotional response.

Classical conditioning: After repeated pairing of a neutral conditioned stimulus (CS), such as a tone, with an unconditioned stimulus (US) that elicits an automatic unconditioned response (UR) to the US, the CS when presented alone elicits the UR and is called a conditioned response (CR) to the CS.

Dopamine: Neuromodulator associated with reward-motivated behavior and motor control. All addictive drugs increase the level of dopamine activity, including cocaine, amphetamine, and methamphetamine. Antipsychotic drugs attenuate dopamine activity. The tremor and motor impairment in Parkinson's disease is caused by loss of dopamine-secreting neurons in the substantia nigra of the basal ganglia.

GABA (γ -aminobutyric acid): Inhibitory neurotransmitter.

Globus pallidus (Pallidum): The output nucleus of the basal ganglia involved in the regulation of voluntary movement.

Hippocampus: Receives converging inputs from associational areas of the cerebral cortex; feedback connections are essential for the consolidation of long-term declarative memories.

Ideal observer/planner: System that performs a perceptual/motor task in an optimal way.

Limbic system: A set of brain structures involved with a variety of basic functions linked to motivation and survival (see Fig. 2). The amygdala in particular is involved in strong emotions such as fear and pleasure and is a gateway between limbic structures and the basal ganglia.

Nucleus accumbens: Part of the basal ganglia, forming the ventral striatum that receives inputs from the prefrontal cortex. Involved in hedonic experiences including laughter and reward, as well as fear, aggression, impulsivity, and addiction.

Operant conditioning: Reinforcement learning in which the consequences of a choice may reinforce or inhibit recurrence of that behavior.

Prefrontal cortex (PFC): The anterior part of the frontal lobes of the brain, lying in front of the motor and premotor areas. The PFC is involved with planning, decision making, and social behavior.

Prospective optimization: Planning future actions to optimize rewards

Orbitofrontal cortex: The part of the prefrontal cortex over the orbit of the eyes that is engaged in evaluating rewards and emotional states in decision making.

Parkinson's disease: A motor system disorder, the result of the loss of dopamine-producing brain cells. The four primary symptoms are tremor (trembling in hands, arms, legs, jaw, and face); rigidity (stiffness of the limbs and trunk); bradykinesia (slowness of movement); and postural instability (impaired balance and coordination).

Reinforcement learning: Area of control theory concerned with how guiding to maximize cumulative reward. Closely related to classical and operant conditioning.

Striatum: Region of the basal ganglia receiving input from the cerebral cortex, involved with coordinating motivation and action. The *Caudate* and *Putamen* are subdivisions.

Substantia nigra pars compacta: Midbrain nucleus containing dopamine neurons that project to the striatum and cerebral cortex. Depletion of dopamine leads to Parkinson's disease.

Temporal-difference (TD) learning: Online reinforcement-learning algorithm based on reward prediction error.

Thalamus. Relays sensory information to the cerebral cortex; receives feedback from the cortex that globally coordinates coherent cortical activity during sleep.

Ventral tegmental area: Origin of the dopaminergic cell bodies that project to the striatum and cerebral cortex and involved in reward cognition, motivation, and drug addiction. ■

REFERENCES

- [1] R. E. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.
- [2] R. A. Rescorla and A. R. Wagner, *A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement*. New York, NY, USA: Appleton-Century-Crofts, 1972.
- [3] R. S. Sutton and A. G. Barto, "Toward a modern theory of adaptive networks: Expectation and prediction," *Psychol. Rev.*, vol. 88, pp. 135–170, Mar. 1981.
- [4] R. S. Sutton, "Temporal credit assignment in reinforcement learning," Ph.D. dissertation, Dept. Comput. Sci., Univ. Massachusetts, Amherst, MA, USA, 1984.
- [5] A. G. Barto, R. S. Sutton, and C. J. C. H. Watkins, "Learning and sequential decision making," *Learning and Computational Neuroscience*. Cambridge, MA, USA: MIT Press, 1990, pp. 539–602.
- [6] T. J. Sejnowski, "Learning optimal strategies in complex environments," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 20 151–20 152, Nov. 23, 2010.
- [7] G. Tesauro, "Temporal difference learning and TD-Gammon," *Commun. ACM*, vol. 38, pp. 58–68, 1995.
- [8] P. Abbeel, A. Coates, and A. Y. Ng, "Autonomous helicopter aerobatics through apprenticeship learning," *Int. J. Robot. Res.*, vol. 29, pp. 1608–1639, 2010.
- [9] T. M. Jessell, E. R. Kandel, and J. H. Schwartz, *Principles of Neural Science*, 4th ed. New York, NY, USA: McGraw-Hill, 2000.
- [10] P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe, *The Hippocampus Book*. Oxford, U.K.: Oxford Univ. Press, 2006.
- [11] P. S. Churchland, V. S. Ramachandran, and T. J. Sejnowski, *A Critique of Pure Vision*. Cambridge, MA, USA: MIT Press, 1994.
- [12] W. S. Geisler, "Sequential ideal-observer analysis of visual discriminations," *Psychol. Rev.*, vol. 96, pp. 267–314, Apr. 1989.
- [13] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 167–192, 2008.
- [14] S. Hecht, S. Shlaer, and M. H. Pirenne, "Energy, quanta, vision," *J. Gen. Physiol.*, vol. 25, pp. 819–840, Jul. 20, 1942.
- [15] H. B. Barlow, "Redundancy reduction revisited," *Network, Comput. Neural Syst.*, vol. 12, pp. 241–253, 2001.
- [16] R. K. Maloney, G. J. Mitchison, and H. B. Barlow, "Limit to the detection of Glass patterns in the presence of noise," *J. Opt. Soc. Amer. A*, vol. 4, pp. 2336–2341, 1987.
- [17] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 10.1–10.26, 2008.
- [18] L. T. Maloney, "Statistical decision theory and biological vision," in *Perception and the Physical World*, D. Heyer and R. Mausfeld, Eds. New York, NY, USA: Wiley, 2002, pp. 145–189.
- [19] E. P. Simoncelli, "Optimal estimation in sensory systems," in *The Cognitive Neurosciences IV*. Cambridge, MA, USA: MIT Press, 2009.
- [20] H. Zhang and L. T. Maloney, "Decision-theoretic models of visual perception and action," *Vis. Res.*, vol. 50, pp. 2362–2374, 2010.
- [21] W. S. Geisler, "Contributions of ideal observer theory to vision research," *Vis. Res.*, vol. 51, pp. 771–781, 2011.
- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2007.

- [23] R. P. Rao, "Bayesian computation in recurrent neural circuits," *Neural Comput.*, vol. 16, pp. 1–38, Jan. 2004.
- [24] R. A. Jacobs, "What determines visual cue reliability?" *Trends Cogn. Sci.*, vol. 6, pp. 345–350, 2002.
- [25] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: In defense of weak fusion," *Vis. Res.*, vol. 35, pp. 389–412, Feb. 1995.
- [26] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, Jan. 24, 2002.
- [27] D. Alais and D. Burr, "The ventriloquist effect results from near-optimal bimodal integration," *Current Biol.*, vol. 14, pp. 257–262, Feb. 2004.
- [28] S. Gepshtein and M. S. Banks, "Viewing geometry determines how vision and haptics combine in size perception," *Current Biol.*, vol. 13, pp. 483–488, Mar. 18, 2003.
- [29] S. Gepshtein, A. Seydell, and J. Trommershauser, "Optimality of human movement under natural variations of visual-motor uncertainty," *J. Vis.*, vol. 7, pp. 13–13-18, 2007.
- [30] D. Gabor, "Theory of communication," *Proc. Inst. Electr. Eng.*, vol. 3, pp. 429–457, 1946.
- [31] J. Feldman, "Bayes and the simplicity principle in perception," *Psychol. Rev.*, vol. 116, pp. 875–887, Oct. 2009.
- [32] H. L. Resnikoff, *The Illusion of Reality: Topics in Information Science*. New York, NY, USA: Springer-Verlag, 1985.
- [33] J. Feldman and M. Singh, "Information along contours and object boundaries," *Psychol. Rev.*, vol. 112, pp. 243–252, Jan. 2005.
- [34] G. Hatfield and W. Epstein, "The status of the minimum principle in the theoretical analysis of visual perception," *Psychol. Bull.*, vol. 97, pp. 155–186, Mar. 1985.
- [35] J. Trommershauser, L. T. Maloney, and M. S. Landy, "Statistical decision theory and the selection of rapid, goal-directed movements," *J. Opt. Soc. Amer. A, Opt. Image Sci. Vis.*, vol. 20, pp. 1419–1433, Jul. 2003.
- [36] J. Trommershauser, S. Gepshtein, L. T. Maloney, M. S. Landy, and M. S. Banks, "Optimal compensation for changes in task-relevant movement variability," *J. Neurosci.*, vol. 25, pp. 7169–7178, Aug. 2005.
- [37] L. Chukoskie, J. Snider, M. C. Mozer, R. J. Krauzlis, and T. J. Sejnowski, "Learning where to look for a hidden target," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. Suppl. 2, pp. 10438–10445, Jun. 2013.
- [38] J. Snider, "Optimal random search for a single hidden target," *Phys. Rev. E*, vol. 83, 2011, DOI: 10.1103/PhysRevLett.112.068103.
- [39] E. A. Schilman, H. B. Uylings, Y. Galis-de Graaf, D. Joel, and H. J. Groenewegen, "The orbital cortex in rats topographically projects to central parts of the Caudate-Putamen complex," *Neurosci. Lett.*, vol. 432, pp. 40–45, Feb. 2008.
- [40] V. B. Tatler B. W., "The prominence of behavioural biases in eye guidance," *Vis. Cogn.*, vol. 17, pp. 1029–1054, 2009.
- [41] R. Shadmehr, J. J. Orban de Xivry, M. Xu-Wilson, and T. Y. Shih, "Temporal discounting of reward and the cost of time in motor control," *J. Neurosci.*, vol. 30, pp. 10 507–10 516, Aug. 2010.
- [42] M. M. Chun and Y. Jiang, "Contextual cueing: Implicit learning and memory of visual context guides spatial attention," *Cogn. Psychol.*, vol. 36, pp. 28–71, Jun. 1998.
- [43] M. Stephenson-Jones, J. Ericsson, B. Robertson, and S. Grillner, "Evolution of the basal ganglia: Dual-output pathways conserved throughout vertebrate phylogeny," *J. Comput. Neurology*, vol. 520, pp. 2957–2973, Sep. 2012.
- [44] S. Ritter and L. Stein, "Self-stimulation in the mesencephalic trajectory of the ventral noradrenergic bundle," *Brain Res.*, vol. 81, pp. 145–157, Nov. 1974.
- [45] P. R. Montague, P. Dayan, and T. J. Sejnowski, "A framework for mesencephalic dopamine systems based on predictive Hebbian learning," *J. Neurosci.*, vol. 16, pp. 1936–1947, Mar. 1996.
- [46] W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," *Science*, vol. 275, pp. 1593–1599, Mar. 1997.
- [47] M. H. Matsumoto and O., "Lateral habenula as a source of negative reward signals in dopamine neurons," *Nature*, vol. 447, pp. 1111–1115, 2007.
- [48] G. Morris, A. Nevet, D. Arkadir, E. Vaadia, and H. Bergman, "Midbrain dopamine neurons encode decisions for future action," *Nature Neurosci.*, vol. 9, pp. 1057–1063, Aug. 2006.
- [49] D. A. Peterson, C. Elliott, D. D. Song, S. Makeig, T. J. Sejnowski, and H. Poizner, "Probabilistic reversal learning is impaired in Parkinson's disease," *Neuroscience*, vol. 163, pp. 1092–1101, Nov. 2009.
- [50] L. Green and J. Myerson, "A discounting framework for choice with delayed and probabilistic rewards," *Psychol. Bull.*, vol. 130, pp. 769–792, Sep. 2004.
- [51] K. Jimura, J. Myerson, J. Hildgard, T. S. Braver, and L. Green, "Are people really more patient than other animals? Evidence from human discounting of real liquid rewards," *Psychol. Bull. Rev.*, vol. 16, pp. 1071–1075, Dec. 2009.
- [52] S. Kobayashi and W. Schultz, "Influence of reward delays on responses of dopamine neurons," *J. Neurosci.*, vol. 28, pp. 7837–7846, Jul. 30, 2008.
- [53] V. Voon and S. H. Fox, "Medication-related impulse control and repetitive behaviors in Parkinson disease," *Archives Neurology*, vol. 64, pp. 1089–1096, Aug. 2007.
- [54] R. Cools, "Dopaminergic modulation of cognitive function—implications for L-DOPA treatment in Parkinson's disease," *Neurosci. Biobehav. Rev.*, vol. 30, pp. 1–23, 2006.
- [55] C. R. Housden, S. S. O'Sullivan, E. M. Joyce, A. J. Lees, and J. P. Roiser, "Intact reward learning but elevated delay discounting in Parkinson's disease patients with impulsive-compulsive spectrum behaviors," *Neuropsychopharmacology*, vol. 35, pp. 2155–2164, Oct. 2010.
- [56] P. Redgrave, M. Rodriguez, Y. Smith, M. C. Rodriguez-Oroz, S. Lehericy, H. Bergman, Y. Agid, M. R. DeLong, and J. A. Obeso, "Goal-directed and habitual control in the basal ganglia: Implications for Parkinson's disease," *Nature Rev. Neurosci.*, vol. 11, pp. 760–772, Nov. 2010.
- [57] H. H. Yin and B. J. Knowlton, "The role of the basal ganglia in habit formation," *Nature Rev. Neurosci.*, vol. 7, pp. 464–476, Jun. 2006.
- [58] R. D. Rogers, A. M. Owen, H. C. Middleton, E. J. Williams, J. D. Pickard, B. J. Sahakian, and T. W. Robbins, "Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex," *J. Neurosci.*, vol. 19, pp. 9029–9038, Oct. 1999.
- [59] H. Nakahara, K. Doya, and O. Hikosaka, "Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences—A computational approach," *J. Cogn. Neurosci.*, vol. 13, pp. 626–647, Jul. 2001.
- [60] K. Semendeferi, A. Lu, N. Schenker, and H. Damasio, "Humans and great apes share a large frontal cortex," *Nature Neurosci.*, vol. 5, pp. 272–276, Mar. 2002.
- [61] N. Parush, N. Tishby, and H. Bergman, "Dopaminergic balance between reward maximization and policy complexity," *Front. Syst. Neurosci.*, vol. 5, 2011, DOI: 10.3389/fnsys.2011.00022.
- [62] R. S. Sutton, "Integrated modeling and control based on reinforcement learning and dynamic programming," in *Advances in Neural Information Processing Systems*, vol. 3, D. S. Touretzky, Ed. Cambridge, MA, USA: MIT Press, 1991, vol. 3, pp. 471–478.
- [63] J. M. Birrell and V. J. Brown, "Medial frontal cortex mediates perceptual attentional set shifting in the rat," *J. Neurosci.*, vol. 20, pp. 4320–4324, Jun. 2000.
- [64] V. J. Brown and E. M. Bowman, "Rodent models of prefrontal cortical function," *Trends Neurosci.*, vol. 25, pp. 340–343, Jul. 2002.
- [65] R. Dias, T. W. Robbins, and A. C. Roberts, "Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: Restriction to novel situations and independence from 'on-line' processing," *J. Neurosci.*, vol. 17, pp. 9285–9297, Dec. 1997.
- [66] R. Dias, T. W. Robbins, and A. C. Roberts, "Primate analogue of the Wisconsin card sorting test: Effects of excitotoxic lesions of the prefrontal cortex in the marmoset," *Behav. Neurosci.*, vol. 110, pp. 872–886, Oct. 1996.
- [67] A. Johnson and A. D. Redish, "Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point," *J. Neurosci.*, vol. 27, pp. 12176–12189, Nov. 2007.
- [68] J. A. Weiler, B. Suchan, and I. Daum, "Foreseeing the future: Occurrence probability of imagined future events modulates hippocampal activation," *Hippocampus*, vol. 20, pp. 685–690, Jun. 2010.
- [69] R. G. Benoit, S. J. Gilbert, and P. W. Burgess, "A neural mechanism mediating the impact of episodic prospection on farsighted decisions," *J. Neurosci.*, vol. 31, pp. 6771–6779, May 2011.
- [70] E. Lalo, S. Thobois, A. Sharott, G. Polo, P. Mertens, A. Pogosyan, and P. Brown, "Patterns of bidirectional communication between cortex and basal ganglia during movement in patients with Parkinson disease," *J. Neurosci.*, vol. 28, pp. 3008–3016, Mar. 2008.
- [71] A. B. L. Tort, M. A. Kramer, C. Thorn, D. J. Gibson, Y. Kubota, A. M. Graybiel, and N. J. Kopell, "Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task," *Proc. Nat. Acad. Sci. USA*, vol. 105, pp. 20 517–20 522, Dec. 2008.
- [72] J. W. Kable and P. W. Glimcher, "An 'as soon as possible' effect in human intertemporal decision making: Behavioral evidence and neural mechanisms," *J. Neurophysiol.*, vol. 103, pp. 2513–2531, May 2010.
- [73] J. W. Kable and P. W. Glimcher, "The neural correlates of subjective value during intertemporal choice," *Nature Neurosci.*, vol. 10, pp. 1625–1633, Dec. 2007.

- [74] P. D. Martin and T. Ono, "Effects of reward anticipation, reward presentation, spatial parameters on the firing of single neurons recorded in the subiculum and nucleus accumbens of freely moving rats," *Behav. Brain Res.*, vol. 116, pp. 23–38, 2000.
- [75] C. M. Pennartz, E. Lee, J. Verheul, P. Lipa, C. A. Barnes, and B. L. McNaughton, "The ventral striatum in off-line processing: Ensemble reactivation during sleep and modulation by hippocampal ripples," *J. Neurosci.*, vol. 24, pp. 6446–6456, 2004.
- [76] J. Peters and C. Buchel, "Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-meditemporal interactions," *Neuron*, vol. 66, pp. 138–148, Apr. 2010.
- [77] D. Hassabis, D. Kumaran, and E. A. Maguire, "Using imagination to understand the neural basis of episodic memory," *J. Neurosci.*, vol. 27, pp. 14365–14374, Dec. 2007.
- [78] K. Gutbrod, C. Krouzel, H. Hofer, R. Muri, W. Perrig, and R. Ptak, "Decision-making in amnesia: Do advantageous decisions require conscious knowledge of previous behavioural choices?" *Neuropsychologia*, vol. 44, pp. 1315–1324, 2006.
- [79] P. Boyer, "Evolutionary economics of mental time travel?" *Trends Cogn. Sci.*, vol. 12, pp. 219–224, Jun. 2008.
- [80] S. J. Mithen, "Key changes in the evolution of human psychology," in *The Evolution of Mind: Fundamental Questions and Controversies*, S. W. Gangestad and J. A. Simpson, Eds. New York, NY, USA: Guilford Press, 2007, pp. 256–268.
- [81] D. J. Foster and M. A. Wilson, "Reverse replay of behavioral sequences in hippocampal place cells during the awake state," *Nature*, vol. 440, pp. 680–683, 2006.
- [82] K. Diba and G. Buzsaki, "Forward and reverse hippocampal place-cell sequences during ripples," *Nature Neurosci.*, vol. 10, pp. 1241–1242, Oct. 2007.
- [83] D. R. Addis, T. Cheng, R. P. Roberts, and D. L. Schacter, "Hippocampal contributions to the episodic simulation of specific and general future events," *Hippocampus*, vol. 21, pp. 1045–1052, Oct. 2011.
- [84] D. L. Schacter, D. R. Addis, and R. L. Buckner, "Remembering the past to imagine the future: The prospective brain," *Nature Rev. Neurosci.*, vol. 8, pp. 657–661, Sep. 2007.
- [85] S. Haykin, *Cognitive Dynamic Systems: Perception-Action Cycle, Radar and Radio*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [86] S. Haykin and J. N. Fuster, "On cognitive dynamic systems: Cognitive neuroscience and engineering learning from each other," *Proc. IEEE*, vol. 102, no. 4, pp. 608–628, Apr. 2014.

ABOUT THE AUTHORS

Terrence J. Sejnowski (Fellow, IEEE) received the B.S. degree in physics from the Case Western Reserve University, Cleveland, OH, USA, and the Ph.D. degree in physics from Princeton University, Princeton, NJ, USA, in 1978.

In 1982, he joined the faculty of the Department of Biophysics, The Johns Hopkins University, Baltimore, MD, USA. He is an Investigator with the Howard Hughes Medical Institute and a Professor at the Salk Institute for Biological Studies, La Jolla, CA, USA, where he directs the Computational Neurobiology Laboratory, and Professor of Biology at the University of California at San Diego, La Jolla, CA, USA. The long-range goal of his research is to build linking principles from brain to behavior using computational models. This goal is being pursued with a combination of theoretical and experimental approaches at several levels of investigation ranging from the biophysical level to the systems level. The issues addressed by this research include how sensory information is represented in the visual cortex.

Dr. Sejnowski is one of only 12 living scientists in all three of the National Academies: Sciences, Engineering, and Medicine. He received the IEEE Neural Networks Pioneer Award in 2002 and the IEEE Frank Rosenblatt Award in 2013.



Gary Lynch received the B.S. degree in psychology from the University of Delaware, Newark, DE, USA, and the Ph.D. degree in psychology from Princeton University, Princeton, NJ, USA, in 1968.

He is currently a Senior Professor in the Department of Psychiatry, University of California at Irvine, Irvine, CA, USA. He has published over 600 papers, been awarded 25 patents, and is recognized by Thomson-Reuters as one of the most highly cited authors in neuroscience. He pioneered the study of long-term potentiation (LTP), a form of synaptic plasticity now regarded as the substrate for memory, and made a series of foundational discoveries concerning the cellular basis of the effect. Related to this, he is the coinventor of the ampakines, a new class of drugs being tested for cognitive enhancement and treatment of diverse psychiatric disorders. He has published two books concerning the evolution of brain size and the related emergence of forebrain networks underlying cognition.



Howard Poizner received the B.A. degree from the University of Texas at Austin, Austin, TX, USA and the M.A. and Ph.D. degrees in cognitive neuroscience from Northeastern University, Boston, MA, USA, in 1977.

He is a Research Scientist in the Institute of Neural Computation, a member of the Program in Neurosciences, the Institute for Engineering in Medicine, and the Kavli Institute for Brain and Mind at the University of California at San Diego, La Jolla, CA, USA. He also is Professor Emeritus of Neuroscience at Rutgers University, New Brunswick, NJ, USA. He was a Postdoctoral Fellow, Staff Scientist, and Associate Director of the Laboratory for Cognitive Neuroscience at the Salk Institute for Biological Studies, La Jolla, CA, USA, from 1978 to 1989. From 1989 to 2004, he was a Professor and then Distinguished Professor of Neuroscience at Rutgers University. He moved to the University of California at San Diego (UCSD), La Jolla, CA, USA, in 2005. His research interests involve unsupervised learning, the neural control of movement, and the integration of brain and movement recordings in virtual environments.

Dr. Poizner is the 2002 recipient of the Rutgers University Board of Trustees Excellence in Research Award.



Sergei Gepshtein received the M.Sc. degree in neurobiology from The Weizmann Institute of Science, Rehovot, Israel, and the Ph.D. degree in psychology from the University of Virginia, Charlottesville, VA, USA, in 2001, where he studied spatiotemporal perceptual organization.

He is a Staff Scientist at the Vision Center Laboratory and a member of the Center for Neurobiology of Vision and Systems Neurobiology Laboratories, Salk Institute for Biological Studies, La Jolla, CA, USA. Following a postdoctoral fellowship in vision science at the University of California at Berkeley, Berkeley, CA, USA, he was a Research Scientist and a member of the Computational Neuroscience Group at RIKEN Brain Science Institute, Saitama, Japan. He joined the Salk Institute in 2007. His research concerns foundations of sensory neuroscience, in particular of visual perception, intersensory integration, and sensorimotor control, with the emphasis of normative models of perception and active behavior.

Dr. Gepshtein is a recipient of the inaugural Hay Grant Award (2013) from the Academy of Neuroscience for Architecture.



Ralph J. Greenspan received the Ph.D. degree in biology from the Brandeis University, Waltham, MA, USA, in 1979.

He has worked on the genetic basis of behavior and brain function in fruit flies (*Drosophila melanogaster*) almost since the inception of the field, studying with one of its founders, Jeffery Hall, at the Brandeis University. He is currently Director of the Center for Brain Activity Mapping, Kavli Institute for Brain and Mind, University of California at San Diego, La Jolla, CA, USA, and Professor in the Division of Biological Science, Neurobiology Section and the Department of Cognitive Science. His research includes studies of the consequences of mutations and localized genetic alterations in the nervous, molecular identification of genes causing naturally occurring variation in behavior, and the genetic analysis of fruit fly sleep and attention. His current research addresses large-scale network interactions pertaining to the action of genes and neurons. In 2011, he was one of the small team of scientists that produced the white paper for the White House Office of Science and Technology Policy that eventuated in the BRAIN Initiative.

