INVITED PAPER

Cognitive Control

Direct information flow in control systems and the concept of information gap, as well as the risks associated with an action or decision policy, are discussed in this paper.

By Simon Haykin, *Life Fellow IEEE*, Mehdi Fatemi, *Student Member IEEE*, Peyman Setoodeh, *Member IEEE*, and Yanbo Xue, *Member IEEE*

ABSTRACT | This paper is inspired by how cognitive control manifests itself in the human brain and does so in a remarkable way. It addresses the many facets involved in the control of directed information flow in a dynamic system, culminating in the notion of information gap, defined as the difference between relevant information (useful part of what is extracted from the incoming measurements) and sufficient information representing the information needed for achieving minimal risk. The notion of information gap leads naturally to how cognitive control can itself be defined. Then, another important idea is described, namely the two-state model, in which one is the system's state and the other is the entropic state that provides an essential metric for quantifying the information gap. The entropic state is computed in the perceptual part (i.e., perceptor) of the dynamic system and sent to the controller directly as feedback information. This feedback information provides the cognitive controller the information needed about the environment and the system to bring reinforcement leaning into play; reinforcement learning (RL), incorporating planning as an integral part, is at the very heart of cognitive control. The stage is now set for a computational experiment, involving cognitive radar wherein the cognitive controller is enabled to control the receiver via the environment. The experiment demonstrates how RL provides the mechanism for improved utilization of computational resources, and yet is able to deliver good performance through the use of planning. The paper finishes with concluding remarks.

KEYWORDS | Cognitive control; cognitive dynamic systems (CDSs); cybernetics; entropic state; information gap; planning; reinforcement learning (RL); two-state model

I. INTRODUCTION

Much has been written on cognitive control in the neuroscience and psychology literature (see, for example, [1] and [2]). In contrast, from an engineering perspective, cognitive control is in its very early stage of development. Looking back on the history of the field of control engineering in the 20th century, we see a trend in the evolution of controllers from simple structures such as open-loop and proportional–integral–derivative (PID) controllers to much more sophisticated ones with features such as optimality, adaptivity, robustness, and intelligence to some extent.

Control systems are usually designed, based on a tradeoff between optimality and robustness. In addition, it is desirable that the controller have the ability to change its behavior in accordance with new circumstances. Adaptive controllers and neurocontrollers have been proposed in the literature to address this issue. In adaptive control, the control problem is formulated in a way that the controller has some adjustable parameters. The controller is also equipped with an adaptation mechanism for updating the parameters according to variations in dynamics of the system with which it interacts as well as the nature of disturbances. Adaptive control systems are inherently nonlinear due to the adaptation mechanism [3]–[6].

While adaptive control is mainly based on parameterized mechanistic modeling, neurocontrollers are based on black-box modeling. Hybrid models (i.e., combination of mechanistic modeling and black-box modeling) can also be used, when finding mechanistic models is straightforward for some parts of the system and difficult for other parts. Control systems can benefit from neural networks in two different ways. One way is to implement the controller using a neural network. In this approach, the controller itself is a neural network; alternatively, the controller may not be a neural network but uses a neural-network-based model of the system under study [7]–[10].

Typically, these controllers function well in structured environments and prespecified conditions, for which they

Manuscript received May 3, 2012; accepted July 9, 2012. Date of publication September 28, 2012; date of current version November 14, 2012. This work was supported by the National Science and Engineering Research Council (NSERC) of Canada.

The authors are with the Cognitive Systems Laboratory, McMaster University, Hamilton, ON L8S 4K1 Canada (e-mail: haykin@mcmaster.ca; mehdi.fatemi@ieee.org; setoodeh@ieee.org; yxue@ieee.org).

Digital Object Identifier: 10.1109/JPROC.2012.2215773

are designed. However, they will not function properly if the system of interest has unmodeled dynamics. In unstructured and/or highly uncertain environments, the presence of a human operator in the control loop is indispensable. In such environments, the controller often reaches points of surprise, for which it has not been programmed. This issue normally arises because the controller is unable to collect the sufficient information it requires to achieve its goals in a self-organized manner. Based on what is known in psychology and neuroscience, it appears that cognition is the needed functionality that should be built into control systems in order to reduce human intervention in the control loop. The article by Buss *et al.* is an endeavor to motivate the need for *cognitive control* in order to elevate the use of automation to the next level [11].

Cognitive control can be viewed as part of a more general framework, called cognitive dynamic systems (CDSs) [12], [13]. The CDS theory is built on Fuster's paradigm of cognition, which states that a cognitive system, in its most general form, has five building blocks, namely, the perception–action cycle (PAC), memory, attention, intelligence, and language [14]. The PAC is the backbone of any closed-loop feedback control system. It can be argued that an adaptive control system may well embody attention and intelligence as well, but lacks memory. Language is more relevant in the context of a network of cognitive agents.

A large percentage of the information processing in the brain is performed in the cortex and it plays a key role in processes attributed to cognition. Regarding the uniform appearance of the cortex, Mountcastle proposed that all regions of the cortex may use a basic information-processing algorithm to accomplish their tasks, regardless of the nature of the information-bearing sensory input. In other words, all kinds of sensory inputs (i.e., visual, auditory, etc.) are coded in a standard form and fed to this basic processing algorithm [15]. Building on Mountcastle's theory, Fuster proposed the concept of *cognit* for knowledge representation in the cerebral cortex [14].

The flow of information in our nervous system plays a critical role in sustaining our vital activities, performing our daily tasks and, even to some extent, determines who we are, especially so when it comes to memory formation. By the same token, the flow of information in man-made machines is of critical importance, regarding performance and robustness of the system. Therefore, controlling the flow of information deserves special attention in the study of cognitive control. Building on achievements of the engineering and neuroscience communities for more than six decades, this paper is aimed at a new generation of control systems that are inspired by the human brain, hence the title: *cognitive control*.

The paper is organized as follows. Section II presents the tale of endeavors on directed information flow in control systems, which has led to the important concept of information gap. It is, in turn, related to the risk associated with an action or decision policy. Having the aim of reducing the information gap, Section III proposes a definition for cognitive control from an engineering perspective, with guidance from neuroscience. Another important notion, namely, the two-state model, is then introduced in Section IV to take account of a quantitative measure of the information gap. Section V describes the reinforcement learning (RL) paradigm and its existence in mammalian brains in order to provide the background for Section VI, where the compositional structure of cognitive control is discussed. Section VII includes two computational experiments on cognitive control in a tracking radar system with emphasis on RL and planning. Finally, Section VIII concludes the paper.

II. CONTROL OF DIRECTED INFORMATION FLOW

This section describes the endeavor of the engineering community to design systems with increased level of sophistication by looking into nature as the main source of inspiration. It is the story of evolution of ideas for more than half a century. It is the tale of *standing on the shoulders of giants* by building on well-established theories, modifying them to extend their applicability to new domains, revisiting them from new perspectives, and integrating them to form more general theoretical frameworks.

Adopting an interdisciplinary approach, after World War II, Wiener had come to the conclusion that the fields of control and communications are both centered around the notion of *information*, where *feedback* plays a key role in information manipulation and decision making [16], [17]. To this end, he came up with the idea of inseparability of these two fields and tried to gather his own work on control and statistical signal processing [18], Shannon's information theory [19], [20], and Kolmogorov's work on prediction theory [21], [22] under a unified umbrella. Wiener called this unifying framework *cybernetics*, which is rooted in the Greek word for *steersman*. As a result of Wiener's close collaboration with the engineer Bigelow and neurophysiologist Rosenblueth, the theory of cybernetics was based on the hypothesis that despite functional differences, machines and living organisms have similar behavioral mechanisms [16], [23]. Wiener also wished to highlight and draw attention to the similarities between the human nervous system on the one hand, and the computation and control in machines, on the other hand, to reach a new interpretation of man, man's knowledge of the universe, and society [16]. In light of these pioneering contributions of Wiener, Dupuy has justifiably argued that cognitive science has its roots in cybernetics [24].

Wiener learned much from the experience gained through working on antiaircraft guns during World War II. In the beginning, human operators were responsible for gun pointing, based on line-of-sight tracking of aircraft. Later on, this human-centered gun-pointing system was replaced with an automatic one by directly coupling a radar to the antiaircraft gun. However, it would still not seem to be practical to completely remove the human operator from the control loop, especially when the behavior of another human being (i.e., the enemy) is needed to be counteracted. By increasing the speed and maneuverability of an aircraft, providing a degree of autonomy for directing the fire was indispensable. However, the system needed to predict the trajectory of the aircraft to make sure that the missile would hit the target at some point of time in the future. Wiener and Bigelow knew that both pilots and gunners would learn their opponent's pattern of behavior and, based on that behavior, improve their own performance over the course of time. To this end, they needed to understand how pilots and gunners were thinking, so as to design a system that would be able to somehow mimic human behavior [16], [23].

Since feedback acts like a double-edged sword, Wiener and Bigelow noticed that as they were pushing for improving the performance of the system, it was possible for the system to become unstable and show oscillatory behavior. They wondered if a similar phenomenon had been observed regarding the nervous system of human beings. In other words, they wondered if there was any nervous-system disorder in which there was no sign of tremor at rest but during an action, the patient was starting to shake more and more severely until he/she could not perform the task. Rosenblueth's answer to this question was *intention tremors* associated with the cerebellum, which is responsible for controlling organized muscular activities. From this pathological resemblance, Wiener, Bigelow, and Rosenblueth concluded that intentional actions in both machines and living organisms can be explained with feedback. They also proposed a behavioral approach for studying systems. This approach is based on an abstract model of the system of interest, which determines the relationship between its input and output. In this abstract model, the output of the system represents any change it causes to the surrounding environment and its input represents the effect of the external events on the system. In this context, feedback provides the means for information manipulation [16], [23].

As mentioned before, Shannon's information theory is one of the pillars of cybernetics. Shannon's information theory was originally developed to mathematically formalize the transmission of signals through a communication channel. The theory provides a quantitative measure of the amount of information, which depends only on the probabilistic structure of the communication channel under study. Information theory has found diverse applications beyond just transmission and compression of data.

Howard emphasized that from a control or decision-making point of view, the probabilistic nature of uncertainties as well as their (economic) impacts on the decision maker must be taken into account and a theory that only deals with probabilities of outcomes may not completely describe the importance of uncertainty to the decision maker. When it comes to allocation of computational resources for information processing, the *value of information* is of critical importance [25].

As Corning stated [26, p. 1], "Shannon's information is blind to the functional properties of the information." According to Corning, the lack of a functional definition of information is the main cause that the full potential of Wiener's cybernetics paradigm [16], [17] has not been realized. Corning suggested the notion of *control information*, which is defined in [26, p. 1] as follows:

> "Control information is not a thing but an attribute of the relationships between things. It is the capacity (*know-how*) to *control the acquisition, disposition, and utilization* of matter/energy in purposive (cybernetic) processes."

Building on this line of thinking, information is the useful or *relevant* portion of the data. Here, usefulness or relevance finds a meaning only in the context of a perceptual task aimed for performing decision making or control [27]. Also, the notion of relevance plays a key role in feature extraction, dimension reduction, and learning. It can be quantified using the concept of *sufficient statistics*, which was proposed by Fisher [28] for parametric distributions. To be more precise, if all the information about the parameters of such distributions can be captured by some functions of a statistical sample, those functions will be considered as sufficient statistics [29]. In this context, the coarsest sufficient partition of random variables, which is drawn from the corresponding distributions, is called *minimal sufficient statistic* [30]. The sufficiency of a statistic for a particular task is specified by the *risk* associated with a control or decision policy. In statistical decision theory, risk is usually defined as the expected loss (or cost) [31].

Hence, the value of data must be related to its complexity after canceling the effect of nuisance factors. Nuisance factors, such as clutter, are the cause of much of the complexity in data. This leads us to the notion of information representation, which is associated with the concept of *Kolmogorov complexity*. Kolmogorov's theory states that the length of an optimal (nonredundant) statement (code) that defines a category is a measure of its complexity [32]. Kolmogorov complexity is closely related to the intuitive notion of conceptual difficulty [33], [34].

Gibson was also one of Shannon's critics; he had a different view on information. Gibson's notion is that information consists of invariants underlying change [35]. Extraction of invariants relates to the explanation of how an observer perceives a true phenomenon of interest, despite uncertain sensory inputs on which the perceptions rely [36]. Inspired by Gibson's work, Soatto called an operational notion of information *actionable information* [27]. Since the question of representation is not quite valid without bringing the task into the equation, he addressed the issue of representation, taking account of decision making and control. Hence, actionable information is a measure for the portion of data that is relevant to the task after removing complexity in the data due to nuisances. In other words, actionable information is defined as the complexity (coding length) of a maximal statistic that is invariant to the nuisances associated with a given task. A statistic (or feature) is invariant if its value does not depend on the nuisance. Maximal invariant is the largest among all invariant statistics in the sense of inclusion of σ-algebras generated by the statistics.[1]

The two attributes of relevance and complexity bring us to the concept of *information bottleneck*, which is closely related to Soatto's approach [38]. Information bottleneck is aimed at finding a compressed, nonparametric, and model-independent representation T of a random variable Y that is as relevant and informative as possible to another random variable X. In this framework, the mutual information[2] between T and Y, $I(Y;T)$, is a measure of complexity, which should be minimized and the mutual information between T and X, denoted by $I(X;T)$, is a measure of informativeness, which should be maximized. Hence, finding the desired representation T can be formulated as an optimization problem in which the tradeoff between complexity and informativeness can be controlled by a Lagrange multiplier. For parametric distributions, minimal sufficient statistics minimize the mutual information $I(Y;T)$ [39]. We may therefore view information bottleneck as a generalization of the classic notion of minimal sufficient statistics [29].

In general, invariant and sufficient statistics may form two different sets; the difference between these two sets leads us to the concept of *information gap*. In order to be able to bridge the information gap, the system must be able to control the perception process [27]. Thus, perception and control are quite intertwined with the emphasis on dependence of perception as a thoughtful activity on the capacity for action. Soatto's approach is tailored for active vision, which deals with a specific type of sensors; the approach is an important step toward a general theory of *controlled sensing* [27].

The concept of controlled sensing is well described by Noë in his book *Action in Perception* [40, p. 1]:

> "*What we perceive* is determined by *what we do* (or what we know how to do); it is determined by *what we are ready to do*. . . . [To be] precise, we *enact* our perceptual experience, we act it out."

Regarding the critical role of information, complex systems can significantly benefit from a mechanism that controls the directed flow of information in a way to decrease a properly defined task-specific information gap. Decreasing the information gap will reduce the risk involved in achieving a satisfactory level of performance. In order to find an appropriate name for such a control mechanism, we may look to the neuroscience literature, in the context of which the term *cognitive control* sounds appealing.

Building on the terminology presented so far, Fig. 1 summarizes the concept of the information gap in a way that is relevant to our context. As illustrated:

- *available information* is extracted from noisy *measurements*, which also includes mapping from measurement space to information space;
- regarding the task at hand, available information can be partitioned into *relevant* and *redundant* information;
- we define *sufficient information* as the required information for performing the task at hand with minimal risk; the mentioned relevant information is therefore the intersection between available information and sufficient information;
- finally, the difference between sufficient information and relevant information forms the *information gap*.

In the following sections, we first look at psychology and neuroscience to pave the way how cognitive control can be defined, and then, we present a systematic way of implementing cognitive control for managing the information gap.

III. HOW DO WE DEFINE COGNITIVE CONTROL?

The term *cognitive control* was first used by psychologists and neuropsychologists. For example, Gardner *et al.* [2] explain six control principles (leveling–sharpening, tolerance for unrealistic experiences, equivalence range, focusing, constricted-flexible control, and field dependence–independence) and 14 experimental tasks to measure them.

Then, Hammond and Summers proposed that [41, p. 1]:

> "Performance in cognitive tasks involves two distinct processes: acquisition of knowledge and cognitive control over knowledge already acquired."

They asserted that acquisition and application of knowledge are independent components of learning in cognitive tasks as well as psychomotor tasks, and then tried to introduce the concept of cognitive control theoretically, and illustrate its empirical significance in studies of human learning, judgment, and interpersonal behavior [41]. They also emphasized the role of task-related feedback as opposed to response-oriented feedback and tried to develop

[1] A nonempty subset of the power set of a nonempty set is a σ-algebra if it includes the empty set and it is closed under complementation and countable unions [37].

[2] Mutual information between two random variables is a measure of the amount of information that one contains about the other. It can also be interpreted as the reduction in the uncertainty about one random variable due to knowledge about the other one [39].

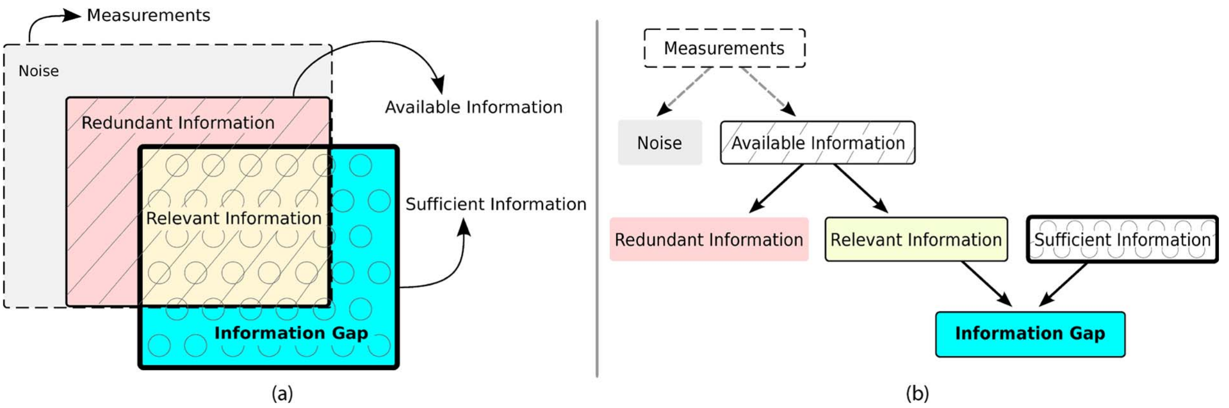


Fig. 1. ***Schematic illustration of the information gap: (a) In this graph, the dashed-line square on the top indicates the noisy measurements, from which the available information is extracted. Dashed line is used to emphasize the fact that measurements may not be in the same space as available information is. In other words, available information is extracted from the measurements by the perception process, and the relationship between them is not necessarily set inclusion. The square at the other corner of the graph demonstrates the sufficient information, which has an overlap with available information. That part is the relevant information, shaded in yellow. The rest of available information, shaded in pink, is therefore redundant information, since it is not relevant to the task at hand. Finally, part of sufficient information, which is left out of available information and shaded in blue, is the information gap. (b) This diagram illustrates the explained concepts in a tree format. The arrows indicate extraction at the top level, inclusion at the middle level, and subtraction at the bottom level.***

a multiple-cue probability learning theory. Some years later, the following definition was proposed in [42, p. 1]:

> "Cognitive control processes refer to our ability to coordinate thoughts and actions in accordance with internal goals."

A similar definition can be found in [43] as well. Yet, another relevant definition presented in [44, p. 1] is as follows:

> "Cognitive control at the neural level is seen as a result of evaluating the probable and actual outcomes of one's actions."

The work done by Feldman and Friston [45] directly relates the neuropsychological ideas to the probabilistic view of an environment. For example, they explain that through *attention*, the brain optimizes its probabilistic representation of the environment. In information theory's terminology, that might be understood to mean a probabilistic representation with minimum entropy due to the fact that entropy is a measure of uncertainty about a random variable [19], [39].

Both in the human brain [46], [47] and in CDSs [12], [13], a *perception* process is performed on sensory measurements. The role of perception is to extract the available information out of *noisy* sensory measurements. In response to information extracted through the perceptor, the human brain performs actions in order to continually enhance this information in subsequent cycles. These actions could be called *cognitive actions*.

For example, say you are in an almost dark room. You might not recognize all the objects clearly. So, the brain will enlarge the pupil size to increase the light entering into the eyes (i.e., to increase information). Suppose the room is too dark so that changing the pupil size does not help. In such a situation, you may perform an external action such as turning on the light. These actions are not being applied to change the state of the environment (for example, the place of objects in the room), but to mitigate the level of uncertainty.

Cognitive Control From an Engineering Perspective: Thus far, the definitions of cognitive control that we have cited, have been drawn from psychology, neuropsychology, and neuroscience. In this paper, we borrow the term cognitive control from neuroscience, and propose the following definition from an engineering perspective with emphasis on controlling the directed flow of information:

> Given a probabilistic dynamic system that embodies the perception-action cycle and, moreover, mimics the human brain, the function of *cognitive control* is to *adapt the directed flow of information* from the perceptual part of the system to its executive part so as to reduce the information gap, which is equivalent to reducing the properly defined risk functional for the task at hand, the reduction being with a probability close to one.

As mentioned before, risk is defined as the expected loss associated with a decision [31]. As a result, there is a requirement for a metric to quantify the information gap. This necessity leads us to the notion of a new type of state to be controlled. This idea is explained in Section IV.

IV. THE TWO-STATE MODEL

At a specific point in time, the state of a dynamic system represents the minimal information that defines the actual condition of the system at that time. By the same token, any change in the state over time (state trajectory) represents the behavior of the system. However, the state is accessible only through noisy measurements, which, in turn, calls for a perception process to provide a *posterior distribution*. As explained in Section II, the difference between the maximal useful information available in the posterior and the sufficient statistics for the given task results in the information gap. This new quantity is thereby defined as the entropic state. The rationale behind choosing this name is that a firsthand candidate for this metric is Shannon's entropy of the posterior due to the fact that entropy can be considered as a global measure of the behavior of the corresponding probability distribution function. This discussion can be summarized in the following two notions of state:

- *system's state*, which is invariant with respect to the measurement process;
- *entropic state*, which is a metric for quantifying the information gap.

Due to uncertainties both in modeling and in measurements, we have to model the state of the system by random variables, and the result of perception will be the posterior distribution, as explained before. The notion of the two mentioned states naturally results in thinking in terms of a *two-state model* of a cognitive-control system, composed as follows:

- *state–space model*, which includes the corresponding mappings from input to system's state and from system's state to output; this model also describes evolution of the system's state over time;
- *entropic-state model* for quantifying the information gap, given the posterior computed by the perception, which depends on environmental uncertainties and disturbances in addition to the sensors' own limitations and modeling errors, as well as the sufficient statistics, which, in turn, depend on the problem under consideration.

Both models may vary from one cycle of the perception–action process to the next in accordance with statistical variations of the environment. Moreover, the feedback information passed on to the cognitive controller is simply the entropic state. As a result, in practice, cognitive control is the paradigm of reducing the entropic state. In Section V, we first present a short review of RL and the fact that it is practiced in mammalian brains, and then we explain that RL is naturally the tool for cognitive control.

V. REINFORCEMENT LEARNING

RL is the mathematical paradigm of learning the best possible action on the sole basis of environmental rewards and punishments (positive and negative rewards). In RL, the goal is to maximize some form of rewards accumulated over the course of time, which are the consequences of a selected action at the current instance. In neuroscience and computational neuroscience, there is now evidence that supports the existence of RL in mammalian brains. This belief has been strongly supported through electrophysiological recordings in behaving animals and functional imaging (fMRI) of human decision-making process [48], [49].

The history of the existence of RL in mammalian brains starts with behavioral studies and goes way back to Pavlovian (classical) conditioning, which involves conditionally learned predictions [50]. Pavlov observed that dogs can be conditioned to predict serving food by a nonrelevant stimulus (conditioning stimulus) such as ringing a bell before they really get served. The dogs then salivate to the ringing of the bell even if there would be no food at all. After the classical conditioning comes the instrumental conditioning, which is learning actions that increase the probability of rewarding events and decrease the probability of adverse events. In other words, instrumental conditioning is a form of learning, in which the behavior is modified by the consequences of actions that result in the behavior. As Niv asserts [48, p. 2]:

> "The study of instrumental conditioning is an inquiry into perhaps the most fundamental form of rational decision-making. This capacity to select actions that influence the environment to one's subjective benefit is the mark of intelligent organisms . . . . [Choosing actions] that will maximize rewards and minimize punishments in an uncertain, often changing, and computationally complex world is by no means a trivial task."

In addition to those behavioral research efforts and perhaps above them, more recent studies have revealed strong neurocellular/molecular evidences of RL. The dopaminergic neurons in the midbrain are now evidently known as the means of performing RL in the brain [48], [49], [51]–[53]. Along the same line of thinking, one of the most important findings, which proves the existence of RL in mammalian brains, is the discovery of a key RL signal in the brain that is understood as the temporal-difference (TD) reward-prediction error [54]–[57]. Additionally, using linear regression, it has been shown that the previously experienced rewards have a part in the dopaminergic response to the current reward, which is exactly according to an exponentially weighted average of past experience, as is implied by the TD learning rule [48], [58], [59].

Computationally, RL theory has been formalized in two parallel but distinct lines of research. In the first line, inspired by Pavlovian (classical) and instrumental conditioning and with the aim of artificial intelligence and agent-based learning, Sutton and Barto shaped the core concepts and algorithms of what is now extensively known as the theory of RL [60]–[64]. In the second line, based on optimal control and Bellman's dynamic programming [65]–[67], Bertsekas and Tsitsiklis developed a group of stochastic approximations, which have been known as neurodynamic programming and approximate dynamic programming [68], [69]. However, it should be noted that, aside from the notations, the difference is mostly due to the definition of reward (in Sutton and Barto's paradigm) and cost (in Bertsekas and Tsitsiklis' paradigm); the former should be maximized, while the latter should be minimized.

There are, on the other hand, several occasions that these (mostly) mathematical theories in the machine-learning literature give insight to computational neuroscientists. For example, inspired by artificial neural networks, Barto and his associates showed that the credit assignment problem[3] can be effectively solved by a learning system, which consists of two neuron-like blocks [61]. One block, called the adaptive critic element (ACE), evaluates different states of the environment, using a TD-like learning rule (from which the TD learning rule was later developed [48]). The other block, called the associative-search element (ASE), then learns to select the best action by means of a trial-and-error process, using the evaluation provided by the first block. Notably [48, p. 9]:

> "These two elements were the precursors of the modern-day Actor/Critic framework for model-free action selection which has been closely associated with reinforcement learning and action selection in the brain."

In fact, the one central idea in the RL literature is TD learning, which is the combination of Monte Carlo methods and dynamic programming ideas [64]. Although TD learning has its early roots in animal psychology and artificial intelligence [71], [72], the first algorithm, called TD(0), and the famous example of *random walk* was created by Sutton [73] (likewise the term TD). Since then, the machine-learning literature has proposed various versions and complementary ideas of the TD learning signal, associated with slightly different model-free RL methods [49], [64]. The two important ideas among them are *Q*-learning and state–action–reward–state–action (SARSA). *Q*-learning is an OFF-policy[4] control algorithm, which was first introduced by Watkins in his Ph.D. dissertation [74], and the convergence proof was later made rigorous by Watkins and Dayan [75]. On the other hand, the SARSA algorithm, which is an ON-policy algorithm, was first explored by Rummery and Niranjan [76], although they called it *modified Q-learning*, and the name SARSA was latter introduced by Sutton [77]. Recent evidence looking primarily at one dopaminergic nucleus seems to support SARSA [78], whereas evidence from a rodent study of the other major dopaminergic nucleus favors *Q*-learning [49], [79]. Whether *Q*-learning is performed in the brain or SARSA (or a combination of them), the fact is that the TD idea is now accepted to be a part of brain's mechanism for selecting the best action.

In what follows, RL and planning will be discussed from a more formal point of view. We will also explain how RL is rationally an intrinsic part of cognitive control and briefly review its attributes. A mathematical treatment of RL in cognitive control has been presented in [80].

VI. COMPOSITIONAL STRUCTURE OF COGNITIVE CONTROL

In this section, we take a closer look at cognitive control in order to provide formal tools for its engineering design. Having the goal of decreasing the information gap and the fact that entropic state, by definition, quantifies the information gap, cognitive control addresses two subproblems:

1) optimal estimation of entropic state;
2) optimal control of entropic state.

In so doing, the perception process is carried out on the sensory measurements, which results in the *posterior* of the system's state. Then, the entropic state is to be calculated from the posterior, and finally it should be controlled in an optimal (or suboptimal) manner. For example, the perception process might be performed by a Bayesian estimator [81], which calculates the posterior of the system's state in each PAC. When the environment is linear with additive white Gaussian noise, the Bayesian filter simplifies to the Kalman filter as a special case [82]. However, when the environment is nonlinear and/or non-Gaussian, the usual procedure is to seek some form of approximation to the Bayesian filter; this approximation may take the form of an extended Kalman filter [83], unscented Kalman filter [84], or cubature Kalman filter (CKF) [85] for nonlinear but Gaussian environments, or a particle filter [86], [87] for general nonlinear and non-Gaussian cases. Then, the entropic state is logically Shannon's entropy [39] of the resulting posterior. Before going further and explaining the configuration of the cognitive controller as an RL agent, let us first take a look at the entire structure of a CDS.

A. Cognitive Control Integrated Inside CDS

Building on Fuster's paradigm, Fig. 2 describes the functional block diagram of a CDS integrating within it: the cognitive controller. In this figure, we readily see that the PAC and memory occupy *physical spaces* of their own. On the other hand, attention and intelligence manifest

[3] The credit assignment problem is the problem of assigning credit or blame for overall outcomes to each of the internal decisions made by the hidden computational units of the distributed learning system [70].

[4] In OFF-policy algorithms, the policy used for learning is different from the one used for selecting control actions, whereas, in ON-policy algorithms, both learning and control share the same policy.

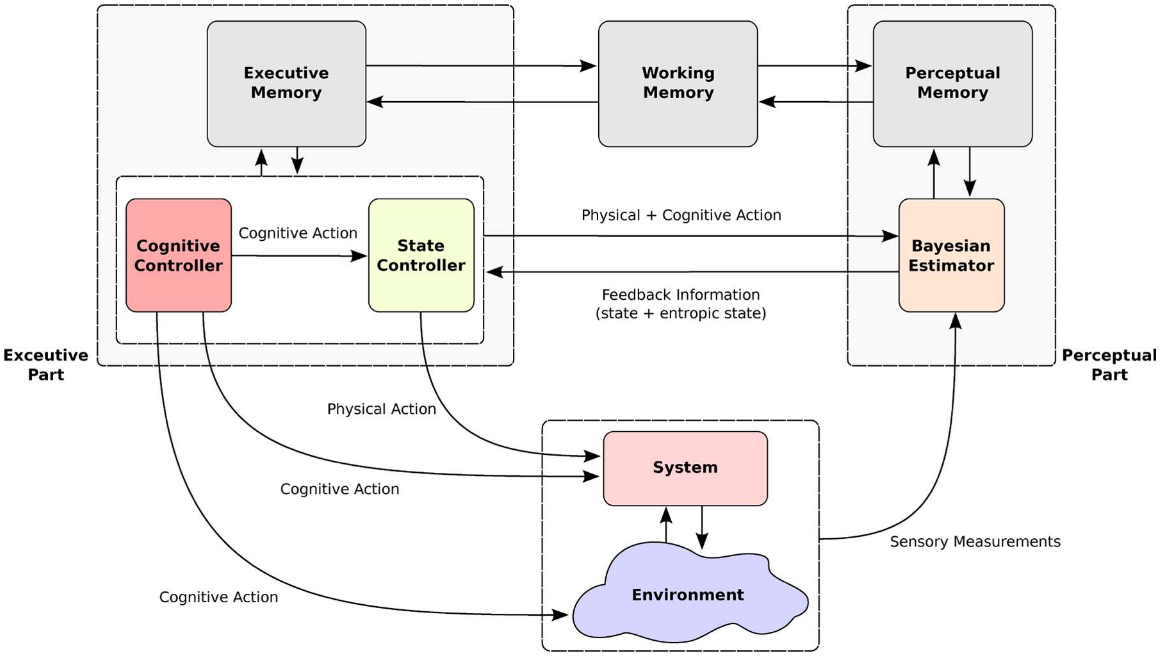


Fig. 2. ***Schematic structure of a cognitive-control system integrated inside a PAC of a CDS, and next to a state controller. It is worth noting that in real-world applications, not all the cognitive action links, shown in the diagram, might necessarily exist. Similarly, in case that the CDS acts as an observer (e.g., cognitive radar systems), the state controller will not be included.***

themselves in the form of *algorithmic mechanisms*, distributed throughout the system.

1) The PAC: Following the terminology of neuroscience, the *perceptual* part of the CDS resides on the right-hand side of the figure, whereas its *executive* counterpart resides on the left-hand side. In effect, the perceptual part of the system, called the *perceptor*, observes the system and the environment directly, whereas the executive part, called the *controller*, observes them indirectly through the "eyes" of the perceptor. This indirect observation of the system and the environment is made feasible by virtue of the *feedback link* that connects the perceptor to the controller.
2) Memory: It builds on the PAC, as depicted in Fig. 2. Specifically, we have:
   - perceptual memory, which is desirably of a hierarchical structure that consists of multiple layers of information processing; the motivation of this hierarchical structure is that of *perceptual abstraction* of the incoming measurements;
   - executive memory, which performs a dual function to the perceptual memory, as shown in Fig. 2; the executive memory has a hierarchical structure of its own;
   - working memory, the function of which is to reciprocally couple the perceptor and controller together, thereby constituting an integrated memory system; this reciprocal coupling makes the cognitive controller operate in a *synchronous fashion* from one PAC to the next.
3) Attention: It manifests itself in an algorithmic manner as perceptual attention in the perceptor and as executive attention in the controller. While perceptual attention deals with the information overflow problem, executive attention implements a version of the *principle of minimum disturbance* [88], [89].
4) Intelligence: It builds on the PAC, memory, and attention, an integrated combination that makes intelligence the most powerful of all the cognitive processes and the most difficult one to define. Similar to attention, intelligence does not occupy a physical place within the CDS, rather its influence is distributed throughout the whole system, and thereby it derives its information-processing power by exploiting all the feedback loops within the CDS, be they global and therefore embodying the environment, or local being confined within the CDS. In short, we may say that the global and local feedbacks are the *facilitator* of computational intelligence in a CDS.

As illustrated in Fig. 2, cognitive actions can influence different parts of the CDS.

- Cognitive actions might be applied to the environment in order to indirectly affect the perception process. An example of this type is turning on the light in a dark room. Here, the physical state includes the position of objects, which is not affected by the light.

- Cognitive actions might also be applied to the system itself in order to reconfigure the sensors and/or actuators, an example of which is changing the pupil size of our eyes according to different light intensities. Another example is changing the transmitted waveforms of a cognitive radar system.
- Additionally, cognitive actions might also be applied as a part of state-control actions (physical actions). In such a case, a physical action is applied to the system, but with the goal of decreasing the information gap (with or without other goals). For instance, consider a quadratic optimal controller with a cost function of the form

$$J = (\mathbf{x} - \mathbf{x}_d)^T \mathbf{Q} (\mathbf{x} - \mathbf{x}_d) + \mathbf{u}^T \mathbf{R} \mathbf{u} \tag{1}$$

to be minimized, where $\mathbf{x}$ and $\mathbf{x}_d$ are the system's state and its corresponding desired-value vectors, $\mathbf{u}$ is the physical control vector, the matrices $\mathbf{Q}$ and $\mathbf{R}$ apply the desired weights for system's state and control, respectively, and the superscript T denotes matrix transposition. To include a cognitive goal, we may add another term to (1) to take care of the information gap as well. The resulting cost function may now be formulated as

$$J = (\mathbf{x} - \mathbf{x}_d)^T \mathbf{Q} (\mathbf{x} - \mathbf{x}_d) + \mathbf{u}^T \mathbf{R} \mathbf{u} + \beta H \tag{2}$$

where H is the entropic state and the scalar β is an importance factor ($\mathbf{Q}$, $\mathbf{R}$, and β are design parameters).

Nevertheless, all these different types of cognitive actions do not necessarily exist in a given problem. In actual fact, a real-world problem might include only one of the aforementioned types of cognitive actions, even without the system's state controller. An example is a cognitive radar system, which only estimates the state of the target without being able to physically control it. In this paper, we mostly focus on cognitive actions, which are directly applied to the system (or to the environment). Such actions call for the implementation of RL and planning in the cognitive-control agent, as discussed next.

B. RL in Cognitive Control

Let us denote the entropic state by $H_{k|k}$,[5] when it reaches the cognitive controller in Fig. 2 at cycle k, after it perceived (estimated) given all the information up to and including cycle k. For optimal control of the entropic state, note that $H_{k|k}$ cannot be controlled directly, even if it gets estimated optimally. Therefore, the primary goal of decreasing $H_{k|k}$ *cannot* be achieved via direct-goal-oriented control techniques, such as full-state feedback control techniques. Additionally, and perhaps more importantly, entropic state is required to be minimized, not just for the next cycle but rather over some look-ahead time horizon. In more formal terms, the cognitive-control action at cycle k should optimally minimize the entropic state at cycle $k+1$ and all the cycles thereafter, based upon knowledge of the environment at cycle k. These two issues naturally form the cognitive-control paradigm as an RL problem. Indeed, RL is at the very heart of a cognitive controller.

In RL, the most basic concept is that of finding a *policy* that is facilitated "only" by *rewards*, which are provided by the environment. Based on the Markov assumption, in RL, we refer to a model by knowing $\mathcal{P}^a_{ss'}$ and $\mathcal{R}^a_{ss'}$, defined, respectively, as follows:

$$\mathcal{P}^a_{ss'} = P[s_{k+1} = s' | s_k = s, a_k = a] \tag{3}$$

and

$$\mathcal{R}^a_{ss'} = E[r_k | s_{k+1} = s', s_k = s, a_k = a] \tag{4}$$

where s is the state supposed to be controlled and a is the action that the control agent can apply to the environment in order to control s. Note that s is not necessarily the physical state of the system. Indeed, in cognitive control, it is the entropic state. $\mathcal{P}^a_{ss'}$ can be found directly from the (stochastic) model that defines the evolution of s over time; however, to find $\mathcal{R}^a_{ss'}$, we need to introduce another equation to model the behavior of the reward function at cycle k (i.e., r_k).

In cognitive control, because the cognitive controller's aim is to decrease the entropic state, a rational reward should include the entropic-state decrement between two subsequent cycles. It is therefore called the *entropic reward*

$$r_k = g_k(H_{k-1|k-1} - H_{k|k}) \tag{5}$$

where $g_k(.)$ is, in general, an arbitrary function.[6] Then, the RL framework ensures decreasing the entropic state not only in the immediate cycle but also in the look-ahead horizon.

To calculate the entropic reward, assuming that the noise distributions in the state–space model can be predicted (or if they are given), then the entropic state can be

[5] Here, $X_{m|n}$ denotes the value of X at time (or at cycle) m, given the information up to and including time (or cycle) n.

[6] However, $g(.)$ should be invertible, as discussed in [80].

predicted using Bayesian paradigm. Therefore, we might benefit from the prediction of future rewards for *planning*.

In RL, there are two distinct but similar concepts as follows [64]:

- *learning* uses actual values of the reward;
- *planning* uses predicted values of the reward.

Planning requires a model of the environment to *simulate* future rewards; however, both learning and planning can use the same algorithm, since they conceptually perform the same task [64]. The important point to note here is that learning can be done only once in each PAC and only for the selected action (since it is based on the actual reward), whereas, in each cycle, planning can be performed for any number of simulated future cycles and for any number of actions. Planning and learning can be integrated to achieve the best result. To this end, Sutton and Barto suggest a simple structure called Dyna [64]. This paradigm can be extended to include the cognitive-control concepts inside the PAC. Details of implementation of RL and planning in cognitive control, however, are beyond the scope of this paper; see [80].

VII. COMPUTATIONAL EXPERIMENT ON COGNITIVE CONTROL

In this section, a target-tracking example is presented to demonstrate cognitive control. We consider the tracking of a falling object with a radar with ten measurements per second, based on the benchmark example presented in [90]. Here, the cognitive actions are "changing" the radar transmitter's waveform parameters in order to mitigate the uncertainty (recall the darkroom example). The target state (i.e., system's state) is $\mathbf{x} = [x_1 x_2 x_3]^T$, where x_1, x_2, and x_3 are the altitude, velocity, and the ballistic coefficient that depends on the target's mass, shape, cross-sectional area, and air density, respectively. In the perceptor, a CKF [85] has been used, which provides the estimated state covariance matrix $P_{k|k}$ at cycle k. Having assumed that the true value of the target's state is required, the information gap will then be a measure that shows how inaccurate the CKF is at each cycle. The entropic state is defined as the Shannon entropy corresponding to the CKF output, and calculated by $H_{k|k} = \det\{P_{k|k}\}$. For the entropic-reward function, we used the following:

$$r_k = \left|\log\left(|H_{k-1|k-1} - H_{k|k}|\right)\right|.\mathrm{sgn}(H_{k-1|k-1} - H_{k|k}) \qquad (6)$$

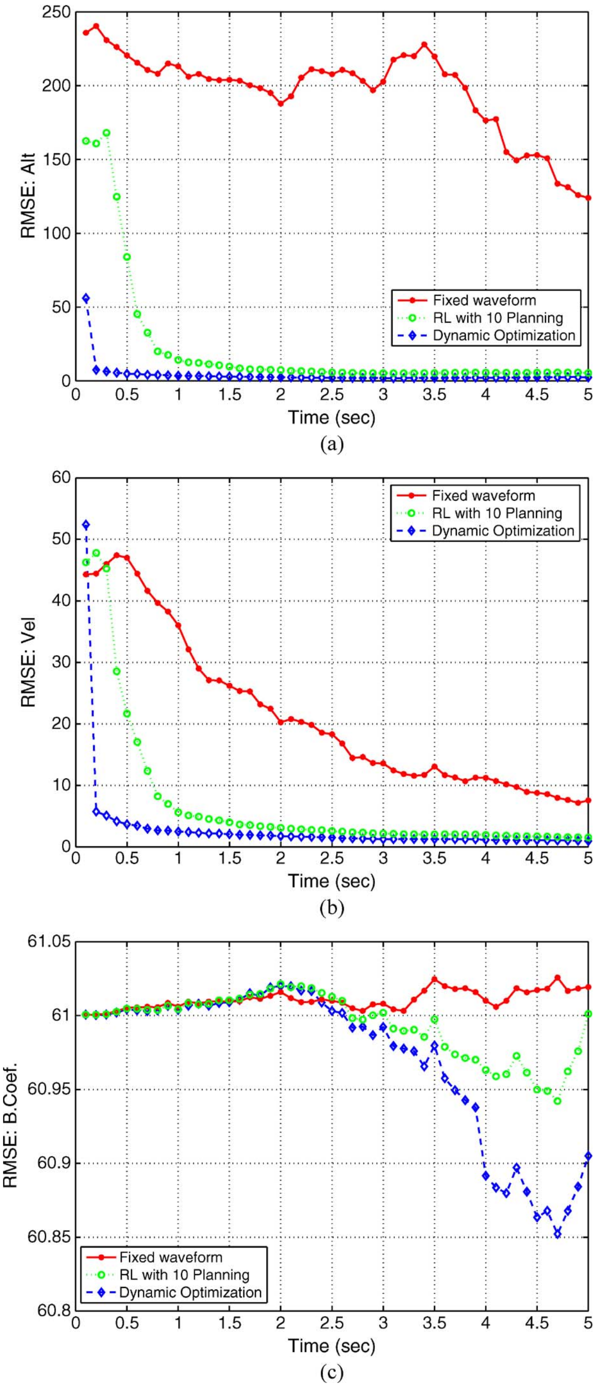
where sgn(.) shows the standard signum function. We have used the logarithm to decrease the intensity of large differences; however, it should be noted that (as can be seen in the results of the next experiment) the difference $|H_{k-1|k-1} - H_{k|k}|$ is never close to zero, so that we have incorrect rewards. In any case, if such events can occur, then $|H_{k-1|k-1} - H_{k|k}|$ should be used instead. This entropic reward also includes a proper sign, which is needed to guide the controller correctly. In the controller side (which is the radar transmitter in this example), there is the possibility of changing the waveform properties in each cycle, which results in 764 cognitive-control actions (i.e., 764 different combinations for the waveform). Applying each action will affect the measurement noise covariance matrix. Finally, Q-learning [64] was chosen as the method of RL for both learning and planning. The emphasis here is on the use of RL and the integration of learning and planning. Therefore, it is assumed that system noise covariance matrix is given and there exists a model for the measurement covariance matrix as a function of control actions [90] (i.e., we do not have entropic-state estimation in this example); details of the implementation have been presented in [80]. All the simulations are performed over 250 Monte Carlo runs to mitigate the effect of randomness. The experiment takes 5 s, therefore, we have 50 PACs.

A. Experiment 1: Suboptimality for Reduced Computational Complexity

In this case study, the functionality of three different radars is compared in terms of their root mean squared error (RMSE). We used the actual value of target state in order to compute RMSE and be able to have the comparison between three different radar configurations. Fig. 3(a)–(c) illustrates the RMSE of the three target state variables, namely, altitude, velocity, and ballistic coefficient, respectively, all of which are plotted versus time. The method of dynamic optimization [90], [91] has been used as a frame of reference, although it may not be used in real time due to its heavy computational load. Additionally, dynamic optimization does not include infinite look-ahead horizon in the sense of Bellman equation [64], [68]. For the cognitive control, here we have Q-learning plus ten actions selected randomly for planning at each PAC. The red bulleted line on the top of the graphs is the radar with no controller (only CKF). The green circled line and blue diamond lines are RL with ten planning and dynamic optimization methods, respectively. The RL method (with ten actions used for planning in each cycle) is almost two orders of magnitude faster than the method using dynamic optimization; hence, RL significantly improves computational complexity at the expense of optimality.

B. Experiment 2: Information-Processing Power of Planning

In Fig. 4, we have illustrated the entropic-state decrement over an increasing number of PACs. The dot magenta line on the top (which almost sticks to the squared-line beneath it) is the fixed-waveform radar, where there are no cognitive-control actions at all. Nevertheless, because CKF is used in the perceptor, the entropic state still decreases (almost two orders of magnitude over the entire 50 PACs). Following that, the blue squared line is for cognitive control only with learning. Since the total number of PACs is far lower than the entire number of possible actions (50 versus 764), this method performs on average, no better than the fixed-waveform method (because *Q*-learning could not converge to any meaningful policy). Then, we retain RL, but this time, we have also added planning. This method is repeated for three different numbers of random actions, which are selected for planning at each cycle: 1) only one random action (red circled line); 2) ten random actions (stared blue line); and 3) 50 random actions (asterisk green line). In the first case that only one action is selected for planning, although one planning is still much lower than the entire number of actions, yet it is enough to demonstrate an obvious improvement. As for the other two cases, they both show more than four orders of magnitude improvement in the entropic-state reduction.



(a)

(b)

(c)

Fig. 3. ***Results of computational experiment of Case Study 1. Figures (a), (b), and (c) illustrate the RMSE for the three state variables correspondingly.***

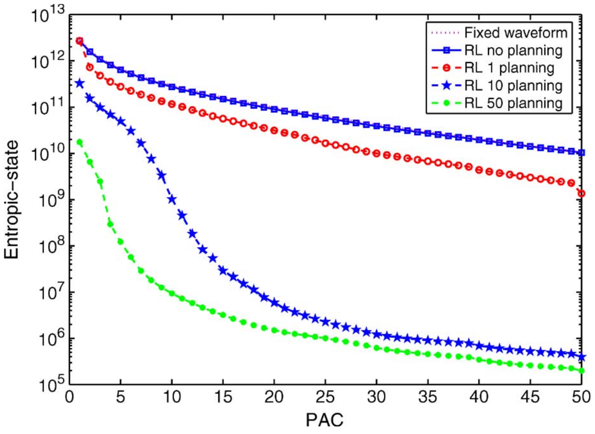


Fig. 4. ***Decreasing the entropic state in a target tracking example using cognitive control. Note that "fixed waveform" and "RL no planning" lines almost coincide with each other on the top of the graph.***

VIII. CONCLUDING REMARKS

A. Summarizing Highlights of the Paper

1) Control of directed information flow in CDS is summed up in the information gap, which is defined as the difference between relevant information (useful part of what is extracted from the measurements) and sufficient information (i.e., the information needed to perform a task of interest with minimal risk).
2) Cognitive control is itself defined as the process of adapting the directed flow of information from the perceptual part of a dynamic system to its executive part, such that the information gap is reduced by an amount equivalent to a reduction in the properly defined risk functional, with a probability close to one.
3) Two-state model, one being the system's state and the other being the entropic state that quantifies the information gap, is defined.
4) RL, exemplified by Q-learning, the employment of which in a cognitive controller is assured by means of the entropic state being computed in the

perceptor and passed directly to the controller as feedback information, is described.

5) Planning, an integral part of RL, requires a model of the environment to simulate future rewards.
6) The following are lessons learned from the computational experiment.
    - The use of RL in a cognitive controller results in a significant reduction in computational resources in exchange for a suboptimal performance.
    - The incorporation of planning into RL enhances the information-processing power of a cognitive controller.

B. Comparison of Cognitive Control Versus Adaptive Control and Neurocontrol

1) Cognitive control versus adaptive control: Adaptation is an integral part of cognition. Therefore, we expect that whatever task is performed by an adaptive controller, the cognitive controller does it better. To elaborate, it can be argued that an adaptive controller could accommodate three of the basic functions of cognition, namely, the PAC, attention, and intelligence. In other words, an adaptive controller lacks memory, whereas memory (and therefore learning) is an integral part of a cognitive controller, hence the ability to outperform an adaptive controller at the expense of increased system complexity.
2) Cognitive control versus neurocontrol: For a neurocontroller, to acquire artificial intelligence and therefore be able to learn from its environment, the traditional approach is to build a neural network into its design. In direct contrast, a cognitive controller looks to neuroscience for guidance. In specific terms, the PAC, memory, and attention are built into the cognitive controller's design; thereby, the controller acquires intelligence, which is the most powerful among all the functions that define cognition. Moreover, the intelligence is distributed throughout the dynamic system via local and global loops. While the neurocontroller works as a whole or it does not work at all [8], cognitive processes (i.e., PAC, memory, attention, and intelligence) can be built into the system in an orderly fashion. Therefore, we expect a cognitive controller to outperform a neurocontroller for a given task, again at the expense of increased complexity.

Most importantly, it should also be emphasized that cognitive control has an intrinsic difference compared to adaptive control and neurocontrol in that the goal of cognitive control is to reduce the information gap. Indeed, as illustrated in Fig. 2, a cognitive-control agent may exist next to or be independent of any other physical controller. In other words, cognitive control is not a *replacement* but an *addition* to a system design paradigm.

To sum up, cognitive control is a new way of thinking about control inspired by the human brain. Over and above the improved utilization of computational resources, yet be able to deliver a good performance through the incorporation of planning in RL, it is in *risk management*, where cognitive control will make its biggest difference to the control literature. ■

Acknowledgment

The detailed feedback notes on different aspects of the paper by four reviewers have not only reshaped the paper into its present form but also made the paper the best it could be.

REFERENCES

[1] R. B. Mars, J. Sallet, M. F. S. Rushworth, and N. Yeung, *Neural Basis of Motivational and Cognitive Control*. Cambridge, MA: MIT Press, 2012.

[2] R. W. Gardner, P. S. Holzman, G. S. Klein, H. P. Linton, and D. P. Spence, "Cognitive control: A study of individual consistencies in cognitive behavior," *Psychol. Issues*, vol. 1, no. 4, pp. 1–186, 1959.

[3] K. J. Åström and B. Wittenmark, *Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[4] I. D. Landau, R. Lozano, M. M'Saad, and A. Karimi, *Adaptive Control: Algorithms, Analysis and Applications*, 2nd ed. New York: Springer-Verlag, 2011.

[5] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness*. Englewood Cliffs, NJ: Prentice-Hall, 1989, reprinted by Dover Publications, 2011.

[6] P. Ioannou and J. Sun, *Robust Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[7] M. Nørgaard, O. Ravn, N. K. Poulsen, and L. K. Hansen, *Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook*. New York: Springer-Verlag, 2000.

[8] T. Hrycej, *Neurocontrol: Towards an Industrial Control Methodology*. New York: Wiley–Interscience, 1997, ser. Adaptive and Learning Systems for Signal Processing, Communications, and Control.

[9] F. W. Lewis, S. Jagannathan, and A. Yesildirak, *Neural Network Control of Robot Manipulators and Non-Linear Systems*. Boca Raton, FL: CRC Press, 1998, ser. Systems and Control.

[10] G. V. Puskorius and L. A. Feldkamp, "Parameter-based Kalman filter training: Theory and implementation," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. New York: Wiley, 2001, pp. 23–67.

[11] M. Buss, S. Hirche, and T. Samad, "Cognitive control," in *The Impact of Control Technology*, T. Samad and A. Annaswamy, Eds. Piscataway, NJ: IEEE Control Systems Society, 2011, pp. 167–173.

[12] S. Haykin, *Cognitive Dynamic Systems*. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2012.

[13] S. Haykin, "Cognitive dynamic systems: Radar, control, and radio," *Proc. IEEE*, vol. 100, no. 7, pp. 2095–2103, Jul. 2012.

[14] J. M. Fuster, *Cortex and Mind: Unifying Cognition*. Oxford, U.K.: Oxford Univ. Press, 2003.

[15] V. B. Mountcastle, *Perceptual Neuroscience: The Cerebral Cortex*. Cambridge, MA: Harvard Univ. Press, 1998.

[16] N. Wiener, *Cybernetics: Or the Control and Communication in the Animal and the Machine*, 2nd. ed. Cambridge, MA: MIT Press, 1965.

[17] N. Wiener, *The Human Use of Human Beings: Cybernetics and Society*. Boston, MA: Houghton Mifflin, 1950.

[18] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. Cambridge, MA: MIT Press, 1964.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948.

[20] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Chicago, IL: Univ. Illinois Press, 1949.

[21] A. N. Shiryayev, *Selected Works of A.N. Kolmogorov: Volume II: Probability Theory and Mathematical Statistics*. New York: Springer-Verlag, 1992.

[22] A. N. Shiryayev, *Selected Works of A. N. Kolmogorov: Volume III: Information Theory and the Theory of Algorithms*. New York: Springer-Verlag, 1992.

[23] R. Seising, *The Fuzzification of Systems*. New York: Springer-Verlag, 2008.

[24] J. P. Dupuy, *On the Origins of Cognitive Science: The Mechanization of the Mind*. Cambridge, MA: MIT Press, 2009.

[25] R. A. Howard, "Information value theory," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-2, no. 1, pp. 22–26, Aug. 1966.

[26] P. Corning, "Control information: The missing element in Norbert Wiener's cybernetic paradigm," *Kybernetics*, vol. 30, no. 9–10, pp. 1272–1288, 2001.

[27] S. Soatto, "Actionable information in vision," Univ. California Los Angeles, Los Angeles, CA, Tech. Rep. CSD090007, 2009.

[28] R. A. Fisher, "On the mathematical foundation of theoretical statistics," *Philosoph. Trans. Roy. Soc. Lond. A*, vol. 222, pp. 309–368, 1922.

[29] O. Shamir, S. Sabato, and N. Tishby, "Learning and generalization with the information bottleneck," *Theor. Comput. Sci.*, vol. 411, no. 29–30, pp. 2696–2711, 2010.

[30] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag, 1998.

[31] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag, 1985.

[32] M. Li and P. M. B. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed. New York: Springer-Verlag, 2008.

[33] J. Feldman, "Minimization of Boolean complexity in human concept learning," *Nature*, vol. 407, pp. 630–633, 2000.

[34] M. Sigman, "Bridging psychology and mathematics: Can the brain understand the brain?" *PLoS Biol.*, vol. 2, no. 9, pp. 1265–1266, 2004.

[35] J. J. Gibson, "The myths of passive perception," *Philosophy Phenomenol. Res.*, vol. 37, no. 2, pp. 234–238, 1976.

[36] J. J. Gibson, *The Ecological Approach to Visual Perception*. New York: Lawrence Erlbaum, 1986.

[37] S. E. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*. New York: Springer-Verlag, 2004.

[38] N. Tishby, "The information bottleneck method," in *Proc. 37th Allerton Conf. Commun. Control Comput.*, 1999, pp. 368–377.

[39] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[40] A. Noë, *Action in Perception*. Cambridge, MA: MIT Press, 2004.

[41] K. R. Hammond and D. A. Summers, "Cognitive control," *Psychol. Rev.*, vol. 79, no. 1, pp. 58–67, 1972.

[42] M. Brass, J. Derrfuss, B. Forstmann, and D. Y. von Cramon, "The role of the inferior frontal junction area in cognitive control," *Trends Cogn. Sci.*, vol. 9, no. 7, pp. 314–316, 2005.

[43] F. Kouneiher, S. Charron, and E. Koechlin, "Motivation and cognitive control in the human prefrontal cortex," *Nature Neurosci.*, vol. 12, no. 7, pp. 939–945, 2009.

[44] W. H. Alexander and J. W. Brown, "Medial prefrontal cortex as an action-outcome predictor," *Nature Neurosci.*, vol. 14, no. 10, pp. 1338–1344, Oct. 2011.

[45] H. Feldman and K. Friston, "Attention, uncertainty, and free-energy," *Frontiers Human Neurosci.*, vol. 4, no. 215, 2010, DOI: 10.3389/fnhum.2010.00215.

[46] R. Rao and D. Ballard, "Dynamic model of visual recognition predicts neural response properties in the visual cortex," *Neural Comput.*, vol. 9, no. 4, pp. 721–763, 1997.

[47] R. Rao and D. Ballard, "Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects," *Nature Neurosci.*, vol. 2, pp. 79–87, 1999.

[48] Y. Niv, "Reinforcement learning in the brain," *J. Math. Psychol.*, vol. 53, *Special Issue on Dynamic Decision Making*, no. 3, pp. 139–154, 2009.

[49] P. Dayan and Y. Niv, "Reinforcement learning: The good, the bad and the ugly," *Current Opinion Neurobiol.*, vol. 18, no. 2, pp. 185–196, Apr. 2008.

[50] R. M. Yerkes and S. Morgulis, "The method of Pavlov in animal psychology," *Psychol. Bull.*, vol. 6, pp. 257–273, 1909.

[51] W. Schultz, "Predictive reward signal of dopamine neurons," *J. Neurophysiol.*, vol. 80, no. 1, pp. 1–27, Jul. 1998.

[52] Y. Niv, N. D. Daw, and P. Dayan, "How fast to work: Response vigor, motivation and tonic dopamine," in *Advances in Neural Information Processing Systems*, vol. 18, Y. Weiss, B. Schölkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, 2005, pp. 1019–1026.

[53] D. J. Surmeier, J. Plotkin, and W. Shen, "Dopamine and synaptic plasticity in dorsal striatal circuits controlling action selection," *Current Opinion Neurobiol.*, vol. 19, pp. 621–628, 2009.

[54] P. R. Montague, P. Dayan, S. J. Nowlan, A. Pouget, and T. J. Sejnowski, "Using aperiodic reinforcement for directed self-organization," in *Advances in Neural Information Processing Systems*, vol. 5, C. L. Giles, S. J. Hanson, and J. D. Cowan, Eds. San Mateo, CA: Morgan Kaufmann, 1993, pp. 969–976.

[55] A. G. Barto, "Adaptive critic and the basal ganglia," in *Models of Information Processing in the Basal Ganglia*, J. C. Houk, J. L. Davis, and D. G. Beiser, Eds. Cambridge, MA: MIT Press, 1995, pp. 215–232.

[56] P. R. Montague, P. Dayan, C. Person, and T. J. Sejnowski, "Bee foraging in uncertain environments using predictive Hebbian learning," *Nature*, vol. 377, pp. 725–728, 1995.

[57] P. R. Montague, P. Dayan, and T. J. Sejnowski, "A framework for mesencephalic dopamine systems based on predictive Hebbian learning," *J. Neurosci.*, vol. 16, no. 5, pp. 1936–1947, 1996.

[58] H. M. Bayer and P. W. Glimcher, "Midbrain dopamine neurons encode a quantitative reward prediction error signal," *Neuron*, vol. 47, no. 1, pp. 129–141, 2005.

[59] H. M. Bayer, B. Lau, and P. W. Glimcher, "Statistics of midbrain dopamine neuron spike trains in the awake primate," *J. Neurophysiol.*, vol. 98, no. 3, pp. 1428–1439, 2007.

[60] R. S. Sutton"A unified theory of expectation in classical and instrumental conditioning," B.S. thesis, Jul. 1978.

[61] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, pp. 834–846, Jul. 1983.

[62] R. S. Sutton, "Temporal credit assignment in reinforcement learning," Ph.D. dissertation, Comput. Sci. Dept., Univ. Massachusetts at Amherst, Amherst, MA, 1984.

[63] R. S. Sutton and A. G. Barto, "Time-derivative models of Pavlovian reinforcement," in *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, M. Gabriel and J. Moore, Eds. Cambridge, MA: MIT Press, 1990, pp. 497–537.

[64] R. S. Sutton and A. G. Barto, *Reinforcement Learning*. Cambridge, MA: MIT Press, 1998.

[65] R. E. Bellman, "Dynamic programming and Lagrange multipliers," *Proc. Nat. Acad. Sci. USA*, vol. 40, no. 10, pp. 767–769, Oct. 1956.

[66] R. E. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

[67] R. E. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, Jul. 1966.

[68] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed., vol. 1, 2. Belmont, MA: Athena Scientific, 2005.

[69] D. P. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[70] S. Haykin, *Neural Networks and Learning Machines*, 3rd. ed. Englewood Cliffs, NJ: Prentice-Hall, 2009.

[71] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Develop.*, vol. 3, no. 3, pp. 210–229, Jul. 1959.

[72] A. H. Klopf, "Brain function and adaptive systems: A heterostatic theory," Air Force Cambridge Res. Lab., Bedford, MA, Tech. Rep. 133, Mar. 1972.

[73] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, Aug. 1988.

[74] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge Univ., Cambridge, U.K., 1989.

[75] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, 1992.

[76] G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," Eng. Dept., Cambridge Univ., Cambridge, U.K., Tech. Rep., Sep. 1994.

[77] R. S. Sutton, "Generalization in reinforcement learning: Successful examples using sparse coarse coding," in *Advances in Neural Information Processing Systems 8*. Cambridge, MA: MIT Press, 1996, pp. 1038–1044.

[78] G. Morris, A. Nevet, D. Arkadir, E. Vaadia, and H. Bergman, "Midbrain dopamine neurons encode decisions for future action," *Nature Neurosci.*, vol. 9, no. 8, pp. 1057–1063, 2006.

[79] M. R. Roesch, D. J. Calu, and G. Schoenbaum, "Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards," *Nature Neurosci.*, vol. 10, no. 12, pp. 1615–1624, 2007.

[80] M. Fatemi and S. Haykin, "On reinforcement learning and planning in cognitive control," *IEEE Trans. Neural Netw. Learn. Syst.*, submitted for publication.

[81] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Autom. Control*, vol. AC-9, no. 4, pp. 333–339, Oct. 1964.

[82] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME/J. Basic Eng.*, vol. 82, *Series D*, pp. 35–45, 1960.

[83] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. New York: Wiley, 2001.

[84] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[85] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[86] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Reading, MA: Artech House, 2004.

[87] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Inst. Electr. Eng. Proc. F—Radar Signal Process.*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[88] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline, and backpropagation," *Proc. IEEE*, vol. 78, no. 9, pp. 1415–1442, Sep. 1990.

[89] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[90] S. Haykin, A. Zia, Y. Xue, and I. Arasaratnam, "Control theoretic approach to tracking radar: First step towards cognition," *Digital Signal Process.*, vol. 21, pp. 576–585, 2011.

[91] S. Haykin, Y. Xue, and P. Setoodeh, "Cognitive radar: Step toward bridging the gap between neuroscience and engineering," *Proc. IEEE*, vol. 100, no. 11, Nov. 2012, DOI: 10.1109/JPROC.2012.2203089.

ABOUT THE AUTHORS

Simon Haykin (Life Fellow, IEEE) received the B.Sc. (first class honors), Ph.D., and D.Sc. degrees in electrical engineering from the University of Birmingham, Birmingham, U.K.

Having worked in signal processing applied to radar and communications for a good part of his professional life, all along in the past 15 years, he had the vision of revisiting the fields of radar and communications from a brand new perspective. That vision became a reality in the early years of this century with the publication of two seminal journal papers: 1) "Cognitive radio: Brain-empowered wireless communications," IEEE Journal on Selected Areas in Communications (February 2005); and 2) "Cognitive radar: A way of the future," IEEE Signal Processing Magazine (January 2006). Cognitive radio and cognitive radar are two important parts of cognitive dynamic systems, on which his new book was published in March 2012. He is the author/coauthor of close to 50 books.

Prof. Haykin is a Fellow of the Royal Society of Canada. In 1999, he was awarded the Honorary Degree of Doctor of Technical Sciences by ETH, Zurich, Switzerland. In 2002, he was the first recipient of the Booker Gold Medal, which was awarded by the International Scientific Radio Union (URSI).

Mehdi Fatemi (Student Member, IEEE) received the B.Sc. degree in physics from Shiraz University, Shiraz, Iran, in 2002 and the M.Sc. degree in computational science (with highest honors) from the Memorial University, St. John's, NL, Canada and the National Research Council Canada, St. John's, NL, Canada, in 2010. He is currently working toward the Ph.D. degree at the Cognitive Systems Laboratory (CSL), McMaster University and the McMaster School of Computational Science and Engineering, McMaster University, Hamilton, ON, Canada.

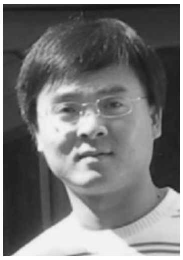
After receiving the M.Sc. degree, he spent six months of full-time work as a Research Engineer in the National Research Council Canada, in collaboration with RAVEN research program. His research interests include cognitive control, complex probabilistic and stochastic systems such as financial systems, as well as computational and mathematical system design.

Mr. Fatemi was named a Fellow of the School of Graduate Studies at the Memorial University, and is a recipient of McMaster graduate scholarship award.

Peyman Setoodeh (Member, IEEE) received the B.Sc. and M.Sc. (first class honor) degrees in electrical engineering from Shiraz University, Shiraz, Iran and the Ph.D. degree in computational engineering and science from McMaster University, Hamilton, ON, Canada.

He is currently a Postdoctoral Fellow at McMaster University, where he collaborates with both the Cognitive Systems Laboratory (CSL) and the Centre for Mechatronics and Hybrid Technology (CMHT) Research. Also, he is a Lecturer in the Department of Electrical and Computer Engineering. His research interests include cognitive machines, complex adaptive systems, nonlinear estimation and control, game theory, and optimization.

Dr. Setoodeh is a recipient of the Monbukagakusho Scholarship from the Ministry of Education, Culture, Sports, Science and Technology in Japan.



Yanbo Xue (Member, IEEE) received the B.Sc. degree in automation engineering and the M.A.Sc. degree in control theory and engineering from Northeastern University, Shenyang, China, in 2001 and 2004, respectively, and the Ph.D. degree in electrical and computer engineering from McMaster University, Hamilton, ON, Canada, in 2010. His Ph.D. dissertation titled "Cognitive radar: Theory and simulations" was the first doctoral dissertation on cognitive radar.

He is currently working as a Postdoctoral Fellow at the McMaster Institute for Automotive Research and Technology (MacAuto). His major research interests are on two aspects: 1) smart grid and hybrid plug-in electric vehicle (HPEV), which includes modeling and optimization of microgrid and smart grid, energy management systems, and vehicle-to-grid (V2G) technology; and 2) cognitive dynamic systems with application to cognitive radar, target detection and tracking, neural networks and learning machines, control theory, and array signal processing.

Dr. Xue was the recipient of Young Scientist Travel Grant (YSTG) of the 2004 International Symposium on Antennas and Propagation (ISAP) held in Japan.