

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/261141210>

# On Cognitive Dynamic Systems: Cognitive Neuroscience and Engineering Learning From Each Other

Article in *Proceedings of the IEEE* · April 2014

DOI: 10.1109/JPROC.2014.2311211

---

CITATIONS

23

---

READS

279

2 authors, including:



Joaquin M Fuster

University of California Los Angeles Medical School

223 PUBLICATIONS 15,179 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



The neurophysiological mechanisms of the major cognitive functions: attention, perception, memory, intelligence and language [View project](#)

# On Cognitive Dynamic Systems: Cognitive Neuroscience and Engineering Learning From Each Other

*In this paper, the authors address recent advances and ongoing challenges in reference to cognitive dynamic systems, embodying cognitive perception, and cognitive control.*

By SIMON HAYKIN, *Fellow IEEE* AND JOAQUÍN M. FUSTER

**ABSTRACT** | Cognitive dynamic systems provide a broadly defined platform, whereby engineering learns from cognitive neuroscience, and by the same token, cognitive neuroscience learns from engineering. The first part of the paper is of a tutorial nature, addressing recent advances in cognitive perception and cognitive control, which are the dual of each other. The study of cognitive perception, viewed from the perspective of Bayesian inference, starts with sparse coding, well known in neuroscience. However, sparse coding could become ill-posed, particularly when the signal-to-noise ratio is low. In such situations, stability is a necessary requirement, which can only be satisfied if there is sufficient information in the observables. To satisfy this requirement, the sparse-coding algorithm is augmented by the addition of information filtering (i.e., a special case of Bayesian filtering). Accordingly, the performance of sparse coding is improved under the influence of perceptual attention. This improvement enhances the cognitive perceptor to separate relevant information from irrelevant information. Next, moving into cognitive control, viewed from the perspective of Bellman's dynamic programming, two ideas are exploited: entropic state of the perceptor, and the definition of reward as an invertible function of two entropic states, namely, the current state and its immediate past value. The net result of building on these two ideas is a modified form of Bellman's dynamic programming, and,

therefore, a new reinforcement learning algorithm, which not only outperforms traditional reinforcement learning algorithms, but also offers some highly desirable properties. Among them is a linear law of computational complexity, which is the best that it could be. The second part of the paper addresses two challenging problems: first, how to mediate between cognitive control and cognitive perception and, second, how to formulate a procedure for risk control. The first problem is resolved by making use of probabilistic reasoning, a branch of probability theory, which leads into the formulation of a probabilistic reasoning machine. With this mediation in place, the conditions for overall system stability are derived, thereby confirming the probabilistic reasoning machine as the overall system stabilizer. The second challenge is risk control, which is by far the most challenging of them all: In the presence of an unexpected disturbance in the environment, risk is brought under control by mimicking the predict and preadapt function, which is considered to be the overarching function in the prefrontal cortex of the brain. To be specific, motor control is expanded by the inclusion of a new preadaptive control mechanism, which involves two different sets of actions: One set is made up of possible actions identified by the policy in the motor control. The other set involves a window of experiences (i.e., optimal actions) gained in the past. In a novel way, by exploiting these two sets, we end up with a preadaptive control mechanism in the form of a closed-loop feedback structure, which brings with it control (executive) attention.

Manuscript received February 16, 2014; accepted March 9, 2014. Date of publication March 21, 2014; date of current version March 25, 2014.

**S. Haykin** is with the Department of Electrical & Computer Engineering, McMaster University, Hamilton, ON L8S 4L8 Canada (e-mail: haykin@mcmaster.ca).

**J. M. Fuster** is with the Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles (UCLA), Los Angeles, CA 90095 USA (e-mail: joaquin@ucla.edu).

Digital Object Identifier: 10.1109/JPROC.2014.2311211

0018-9219 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

**KEYWORDS** | Attention; cognition; control; entropic state; intelligence; memory; perception; pre-adaptation; prediction; risk control; self-organization; system stability

## I. INTRODUCTION

### A. Historical Background

Strictly speaking, cognitive neuroscience preceded by more than a century, the emergence of information theory, artificial intelligence, and neural engineering. The origin of cognitive neuroscience can be traced back to the discovery of two areas of the cerebral cortex, Broca [1] and Wernicke [2], who specialized in the articulation and understanding of speech, respectively. The consequent inference of two separate cortical modules for the processing of language led to a search for comparable modules dedicated to other cognitive functions, such as memory, perception, and intelligence. The trend was fostered in animal and human neuropsychology by the discovery of cortical areas, whose lesion led to disorders in visual, auditory, or tactile discrimination. Then, most prominent was the memory disorder caused by lesions of the hippocampus, a parcel of ancient cortex.

The modular model for the study of cognitive neuroscience has continued almost unabated to the present, reinforced by the electrophysiological analysis of sensory areas. This analysis reveals apparent cortical hierarchies of modules, dedicated to progressively more abstract aspects of perception or memory. Since the middle of the last century, however, the modular model has shown progressive signs of strain, while the network model of cognition increasingly displaces it. The latter is also hierarchical, but in it the self-organized unit of memory or knowledge is not the module but the network. The pressure for change came in mid-century from two converging fields.

- 1) Clinical neurology: Geschwind [3] demonstrated that certain cognitive disorders resulted not so much from the damage of specific cortical areas, but from the connections between them. This observation led to a proliferation of neuroanatomical studies of cortical connectivity, providing evidence of widely distributed cortical networks to serve cognition.
- 2) Theoretical and computational approaches to neural networks [4]–[7]: Connectionism is one such theoretical approach that is most germane to current views on cortical cognition, as both of them postulate a complex system of distributed, overlapping, and hierarchically organized networks of knowledge. The connectionism approach shares many common elements with the cognitive cortical paradigm postulated in [8] and [9]. Other contributions to neural networks include novel ideas described in [10] and [11].

In summary, beginning with the 1950s, there has been a continuous, though somewhat muted, dialog between the neuroscientists of cognitive networks and the scientists dedicated to the study of the dynamics of information networks in complex adaptive systems. This paper is an attempt to make that dialog more explicit, with cognitive dynamic system as a new way of thinking.

The idea of cognitive dynamic systems, from an engineering perspective, was inspired by the human brain; its origin is summarized as follows.

- 1) In a predictive Point-of-View article, published in the 2006 PROCEEDINGS OF THE IEEE [12], the idea of cognitive dynamic systems was motivated by two preceding papers: the classic 2005 paper on cognitive radio [13], and the seminal paper written on cognitive radar in 2006 [14].
- 2) In the context of cognitive dynamic systems, the first book to be written on the subject, was published in 2012 [15].
- 3) Then, there was a second predictive Point-of-View article on cognitive dynamic systems that was published in the PROCEEDINGS OF THE IEEE in 2012, where the scope was broadened to include radar, control, and radio [16].

### B. Principles of Human Cognition

In the book entitled *Cortex and Mind: Unifying Cognition* [17], the basic principles of human cognition<sup>1</sup> are identified; they are briefly described in what follows.

- 1) Perception–action cycle. Environmental observables (measurements), coming into the perception part of the brain, are processed to extract relevant information about the environment. This processing continues from one cycle to the next, until a point is reached where any further information gain about the environment becomes essentially too small to be of practical value, assuming that the environment is locally stationary.
- 2) Multilayered memory. Basically, memory consists of three parts:
  - perceptual memory that builds on the perception–action cycle for the extraction of relevant information about the environment;
  - executive memory that builds on feedback information about the environment to produce actions on the environment;
  - perceptual memory and executive memory are reciprocally coupled via working memory.
 Simply put, the function of memory is to predict the consequences of action taken on the environment by the executive part of the brain. In other words, memory builds on the perception–action cycle.
- 3) Attention. Whereas the perception–action cycle and memory occupy distinct physical places in the brain, attention is algorithmic in nature. Specifically, there is perceptual attention in the perception part of the brain, and executive attention in its executive part. Simply put, the function of attention is the efficient allocation and

<sup>1</sup>In actual fact, there is a fifth principle, namely, language; for the present, language has been put aside for another day.

management of resources in the brain. With this function in mind, attention builds on memory and, not only that, but also the perception–action cycle. It is also important to note that feedback (of the negative kind) is the facilitator of attention.

- 4) Intelligence. Building on attention (and, therefore, memory as well as the perception–action cycle), intelligence is the most powerful of all the principles in human cognition, and, therefore, it is difficult to define. Nevertheless, the primary objective of intelligence is optimal control of a target of interest, and doing so in the most effective and efficient manner possible, followed by decision making for action on the environment.

### C. Organization

The rest of the paper, consisting of eight sections, is organized as follows.

Section II describes the diagrammatic structural composition of a cognitive dynamic system that mimics the brain, with emphasis on two topics:

- three kinds of perception–action cycles, whose formulations depend on where they are located within the system;
- the hierarchical structure of memory, the purpose of which is to trade off time for space.

Next, recognizing that the Bayesian approach is necessary to resolve the inherent ambiguity (ill-posed nature) of perceptual inverse problems, be that from a cognitive neuroscience or engineering perspective, Section III covers the following topics:

- Bayesian inference;
- probabilistic modeling;
- statistical analysis;
- maximum *a posteriori* (MAP) rule for parameter estimation.

At this point, the stage is set for how perception in a cognitive dynamic system is modeled. To this end, Section IV on cognitive perception builds on sparse coding, which is well known in cognitive neuroscience. Unfortunately, there are situations encountered in practice, where sparse coding violates the three conditions of Hadamard, namely, existence, uniqueness, and stability. It is the latter condition, where sparse coding would have to be improved under the influence of perceptual attention. To satisfy this requirement, the use of an information filter is put in place, which is a special form of the optimal Bayesian filter.

Just as Bayesian inference is basic to cognitive perception, so it is that Bellman’s dynamic programming underlies cognitive control. To satisfy this latter need, Section V is devoted to Bellman’s dynamic programming for a stochastic environment. The material covered in Section V includes the following topics:

- Markov decision processes;
- Bellman’s optimality equation.

With this material in place, the stage is set for cognitive control, which is discussed in Section VI. The following topics are covered therein:

- the imperfect state-information problem;
- the entropic state model for the perceptor that builds on Shannon’s entropy;
- revisiting Bellman’s optimality equation, with a special focus on the novel notion of entropic state that has a profound importance on the optimal control formulation;
- exploiting another idea, namely, the functional dependence of reward on states;
- explore–exploit tradeoff.

Section VII stresses the idea of a probabilistic reasoning machine that addresses the following topics:

- expansion of entropic states to include the perceptor as well as the controller;
- probabilistic reasoning, well known in probability theory, which acts as the mediator between the perceptor and the controller, thereby providing the basis for overall system stability and self-organization.

Risk management is perhaps the most challenging problem to resolve not just in engineering, but also in cognitive neuroscience. Section VIII addresses this issue by exploiting an overarching preadaptation function in the prefrontal cortex, which builds on the following idea:

“For a prediction in the future to exist, there has to be a past.” Finally, Section IX concludes the paper.

## II. COGNITIVE DYNAMIC SYSTEMS

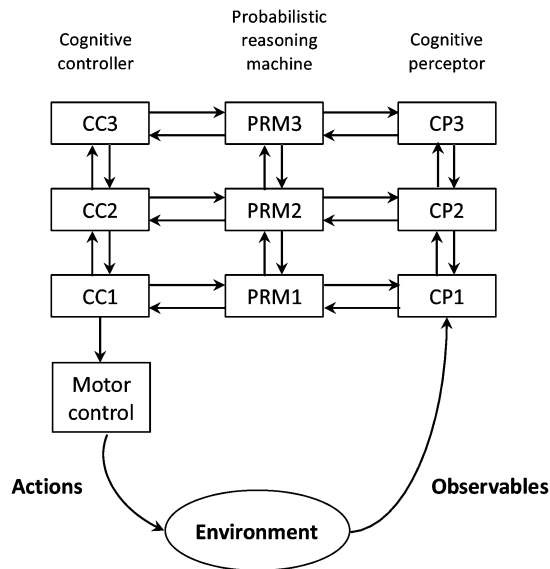
### A. Structural Composition of a Cognitive Dynamic System

Fig. 1 shows the block diagram of a cognitive dynamic system, which, from an engineering perspective, is configured to mimic the brain. In physical terms, the figure closely depicts the two principles of cognition: perception and memory, which are individually discussed in what follows.

1) *Perception–Action Cycles*: The immediate impression we get from the examination of Fig. 1 is the fact that the cognitive dynamic system is a multiple closed-loop feedback system, in which feedback, distributed throughout the system, plays a critical role in how it perceives the world (environment).

The part of the system to the right of the figure is called the cognitive perceptor, and the part of the system to the left of the figure is called the cognitive controller<sup>2</sup>; these

<sup>2</sup>In cognitive neuroscience, the cognitive perceptor is referred to as the sensory system, and the cognitive controller is referred to as the executive system. The principal functional block that reciprocally couples them together is called the working memory.



**Fig. 1. Functional block diagram of hierarchical cognitive dynamic system. CP: cognitive perceptor; CC: cognitive controller; PRM: probabilistic reasoning machine.**

two parts of the cognitive dynamic system are reciprocally coupled by the probabilistic reasoning machine. The global perception–action cycle, so called because it embodies the environment inside its feedback loop, proceeds as follows.

- 1) The cognitive perceptor processes the incoming environmental observables to extract relevant information about the environment. Moreover, it computes feedback information, based on perception errors.
- 2) The feedback information is passed by the cognitive perceptor to the cognitive controller, thereby linking them for the cognitive dynamic system to operate as a whole.
- 3) In response to the feedback information as its input, the cognitive controller acts on the environment, thereby closing the global perception–action cycle.
- 4) The resulting action produces further changes in the environment that lead to new perception and, in this manner, the cycle continues on, until a prescribed goal is realized.

Indeed, it is through the continuation of the global perception–action cycle across time that a cognitive dynamic system acquires a cardinal property, which enables it to continually adapt to changes in the environment by making its own successive changes through experience learned from interactions with the environment. To reemphasize the importance of this capability, we say that the coordination of functions involved in global perception–action cycles across time is a

distinctive characteristic of cognition; hence, the three key roles attributed to time [17]:

- 1) time separates the observables from one cycle to another, so as to guide the overall behavior of the cognitive dynamic system;
  - 2) time separates the observables for perception of the environment from action taken on the environment;
  - 3) time separates feedback information from action.
- The implication of these three key roles of time is profound, prompting us to make the following statement:

“Temporal organization of the overall behavior of the cognitive dynamic system requires the coordination of time: percepts with percepts, actions with actions, and percepts with actions.”

The discussion thus far has focused on the global perception–action cycle. To be specific, there is another kind of cycle called the internally composite perception–action cycle, which distinguishes itself from its global counterpart as follows: Internally composite perception–action cycles embody the cognitive perceptor, the probabilistic reasoning machine, and the cognitive controller inside their feedback loops, completely eschewing the environment. It is this particular perception–action cycle that is responsible for both control attention and perceptual attention.

Finally, we come to the third kind of cycles, called local perception–action cycles. The term local is meant to emphasize the fact that these cycles distinguish themselves on the following account: They are localized inside the cognitive perceptor or cognitive controller.

The practical importance of local perception–actions cycles is summarized as follows [17], [18]:

“The cycles localized in the cognitive perceptor are responsible for perceptual attention and those localized in the cognitive controller are responsible for control (executive) attention.”

Indeed, it is on account of this statement that attention is said to be algorithmic in nature.

## B. Hierarchical Structure of Memory

As the name would imply, perceptual memory is an integral part of the cognitive perceptor. To be more specific, perceptual memory provides the cognitive perceptor with the ability to interpret the observables, so as to distinguish their characteristic features learned in a statistical sense. The learning is conducted through an adaptive matching process, where features in one layer of the cognitive dynamic system are matched to those features computed from the lower layer. Speaking of layers, in a hierarchical perceptual memory in the cognitive perceptor, the observables are processed, unit

by unit.<sup>3</sup> Accordingly, the perceptual constancy across the hierarchical structure of the memory increases in abstraction as we go up from one layer to the next. Just as the perceptual memory resides in the cognitive perceptor, the executive memory resides in the cognitive controller. Contents of the executive memory are continually changed from one global perception–action cycle to the next, as a result of actions performed on the environment by the cognitive controller. In other words, knowledge gained in the executive memory is of an experiential kind; it is updated through actions on the environment that are taken in response to feedback information sent to the cognitive controller by the cognitive perceptor. In moving down across the cognitive controller layer by layer, experiential knowledge becomes increasingly focused on the prescribed goal of interest. In contrasting the perceptual memory with the executive memory, we speak of the following [17]:

- perceptual cognits represent knowledge (i.e., features) contained in the environmental observables;
- executive cognits represent knowledge gained from experience through interactions with the environment.

Quoting from [17]:

“A cognit is an item of knowledge about the world, the self, or the relations between them. Its network structure is made up of elementary representations of perception or action that have been associated with one another by learning or past experience.”

### III. BAYESIAN INFERENCE

Before discussing cognitive perception in the next section, it is instructive to digress briefly in order to present some preparatory material on Bayesian inference, which plays a key role in the study of perception, be that in neuroscience or engineering.

To begin the discussion of Bayesian inference, the term inference is said to be a compact way of referring to the statistical evaluation of a model that is of particular interest. The underlying philosophy of Bayesian inference is then summed up in the following statement, taken in verbatim from [19]:

“A Bayesian approach to a problem starts with the formulation of a model that we hope is adequate to describe the situation of interest. We then formulate a prior distribution over the unknown parameters of the model, which is meant to capture our beliefs about the situation before seeing the data. After observing some data, we apply Bayes rule to obtain a

posterior distribution for these unknowns, which takes account of the prior and the data. From the posterior distribution we can compute the predictive distribution for future observations.”

To proceed with the discussion of Bayesian inference, there are two finite-dimensional spaces.

- 1) A parameter space and an observation space, where the parameter space is hidden from the observer that only has access to the observation space. Therefore, insofar as the observer is concerned, the model parameters are unknown. Let  $\theta$  denote the parameter vector drawn from the parameter space.
- 2) The parameter vector  $\theta$  is mapped onto the observation space by a probabilistic transition mechanism from the parameter space to produce a sample observation vector  $x$ .

We may now make the following two statements that are the dual of each other [20]:

- 1) probabilistic modeling, the aim of which is to formulate the conditional distribution  $p(x|\theta)$  that provides an adequate description of the underlying physical behavior of the observation space;
- 2) statistical analysis, the aim of which is to produce the inverse of probabilistic modeling; this inversion is denoted by the posterior distribution  $p(\theta|x)$ .

To proceed from the conditional distribution to the posterior distribution, we involve Bayes rule, obtaining

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \quad (1)$$

where  $p(\theta)$  is the prior, and  $p(x)$  is the evidence.

*The MAP Rule:* The inversion aspect of the statistical analysis manifests itself in the notion of the likelihood function denoted by  $\ell(\theta|x)$ , for which we introduce the following definition:

$$\ell(\theta|x) = p(x|\theta). \quad (2)$$

Assuming that the parameter vector  $\theta$  is continuous, we may then compute an estimate of the unknown  $\theta$  using the product  $\ell(\theta|x)p(\theta)$ , hence the following statement:

“The MAP estimate, denoted by  $\hat{\theta}_{\text{MAP}}$ , is defined by the maximum value of the product composed of the likelihood function  $\ell(\theta|x)$  and the prior  $p(\theta)$ .”

What is important to note here is the fact that, from an information-theoretic point of view, the MAP estimate is

<sup>3</sup>The term layer applies to the reciprocally coupled perceptor and controller in the cognitive dynamic system, whereas the term unit applies to a component of the perceptor or that of the controller viewed by itself.



absolutely optimum, in that there is no other estimate that can outperform it. However, this unique capability is attained at having to have a prior. Note also that in making this statement, the evidence  $p(x)$  acts as a normalizer and, therefore, has no role to play in the MAP estimate.

If, for some reason, the prior  $p(\theta)$  is not available or difficult to formulate, the MAP estimate is simplified by settling for the maximum likelihood (ML) estimate, where simplicity is gained at the expense of an estimate that is no longer optimal.

## IV. COGNITIVE PERCEPTION

### A. Sparse Coding

To explain what we mean by sparse coding, we can do no better than the following statement, reproduced verbatim from the paper by Olshausen and Field [21]:

“The principle of sparse coding refers to a neural code, in which each sensory input of a nervous system is represented by the strong activation of a relatively small number of neurons out of a large population of neurons in the system.”

In practical terms, the advantages of sparse coding are summarized as follows.

- 1) Storage capacity of the multilayered perceptual memory is increased significantly.
- 2) At the top layer of the perceptual memory, the observables are represented in a manner that are easy to recognize on account of two facts. First, the relevant features of the observables are more likely to be largely separable from the irrelevant features. Second, the relevant features are more stable (robust) in the presence of additive noise.
- 3) Last, energy is conserved and computational costs are minimized.

However, to realize these advantages, we are faced with a sparse-coding problem that is difficult to solve. Thus, setting the advantages of sparse coding just mentioned versus the complexity of how to solve the problem, it reminds us of the no free lunch theorem [22], which, in effect, states that for every gain made, there is a price to be paid.

### B. Ill-Posedness of the Sparse-Coding Problem

From a computational perspective, sparse coding could be viewed as a linear inverse problem in the following twofold manner:

- sparse coding is an inverse problem in the sense that the neural code (i.e., the set of features representing the observables) is hidden from view of the perceptual memory;
- linearization of the problem is proposed to make it mathematically tractable.

Accordingly, the first step in solving the sparse-coding problem is to introduce the following two definitions.

- 1) A neural code, denoted by  $z$ , which defines the feature vector that represents the observables; hence, the neural code is also referred to as the feature vector.
- 2) A dictionary, denoted by a rectangular matrix  $W$ , which consists of generating elements (e.g., Gabor wavelets) that are chosen to be nonorthogonal with respect to each other; the number of rows in the dictionary is much larger than the number of columns.

Using vector  $x$  to denote the observables, we may then write

$$Wz = x \quad (3)$$

which represents a system of linear equations. Because the number of equations is smaller than the number of observables, the system of equations is underdetermined and, therefore, nonsolvable. Turning next to the issue of ill-posedness, a problem is said to be ill-posed if it violates the three Hadamard conditions of well-posedness, namely [23]:

- 1) a solution to the problem exists;
- 2) the solution is unique;
- 3) moreover, the unique solution is stable (robust).

In its basic form, sparse coding violates condition 1) because the sparse-coding problem is underdetermined. To satisfy conditions 1) and 2), the typical approach has been to invoke Tikhonov's regularization theory and extensions thereof [23], [24].

However, regularization by itself may not always be enough to take care of condition 3), that is, stability.

To elaborate on what we have said thus far: When the level of noise power in the observables is low (i.e., the signal-to-noise ratio is high), then the use of regularization to satisfy conditions 1) and 2) could suffice. On the other hand, when the level of noise power is high (i.e., the signal-to-noise ratio is low), condition 3) is vulnerable and can, therefore, be violated. In such a situation, the regularized sparse-coding algorithm can be stabilized by the addition of information filtering that introduces perceptual attention [25]. In the final analysis, we may make the following statement:

“For the sparse-coding algorithm to be well-posed, there would have to be sufficient information in the observables, subject to the provision that the signal-to-noise ratio is not too low.”

There is an interesting convergence here between predictive coding formulations of Bayesian or information filtering and the adaptive response to fluctuations in signal-to-noise ratio. It has recently been proposed that attention can be understood as optimizing the precision

(inverse variance) associated with noise [26]. Practically, this involves the use of precision weighted prediction errors, where precision is encoded by the Kalman gain matrix.<sup>4</sup> If we associate signal to noise with precision, there is a clear link between cognitive attention and perceptual inferences in which we have greater confidence. Furthermore, the Fisher information, to be discussed in the next section, is the conditional precision (inverse variance) under linear (and Gaussian) assumptions. Again, we return to the notion of the inverse covariance or precision as an integral part of information filtering, which can select informative observations or prediction errors for Bayesian updating.

### C. Information filtering: A Special Form of Bayesian Filtering

Consider an environment, the state-space model of which is described by the following pair of related equations:

- 1) process equation, which describes the evolution of the state of the environment across time, with the additive process noise acting as the driving force;
- 2) measurement equation, which describes the dependence of measurements (observables) on the state in the additive presence of measurement noise.

This state-space model (derived from physical considerations) satisfies the Markov property, which means that formulation of the model requires the current state and its immediate past value. This property means that if the initial condition is known, then the next value of the state can be computed; given the value of the state just computed, the second value of the state can be computed, and so it goes on. With the state of the environment being hidden, there are two assumptions that are often made to keep the computational complexity manageable.

- 1) the state-space model is linear;
- 2) the process noise and the measurement noise are statistically independent and they both have (a potentially) different Gaussian distribution.

It turns out that the Bayesian filter is the optimal solution for the pair of equations described by the state-space model. Moreover, the Bayesian filter operates iteratively, with each iteration consisting of two steps: innovation followed by prediction.

The Kalman filter and the information filter are two special cases of the Bayesian filter. They are mathematically equivalent, but their innovation and prediction steps are entirely different. To be more specific:

- 1) the Kalman filter [23] updates the state estimate by propagating the covariance matrix of the state-estimation error vector from one iteration to the next, where this error is defined by the difference between the pseudoactual value of the state (computed from the state-space model) and its estimated value;

- 2) in direct contrast, the information filter [23] updates the information vector (i.e., the counterpart of the state vector in the Kalman filter) by propagating the inverse covariance matrix of the estimation error vector from one iteration to the next. What is important to note here is the fact that the inverse covariance matrix, just defined, is the Fisher information, hence the name of the filter.

Simply put, the Kalman filter ignores the information vector, whereas the information filter ignores the state vector. In any event, the information filter has an advantage over the Kalman filter: the information filter could be initialized with a diffuse prior (i.e., zero information), which simplifies algorithmic formulation of the information filter.

Finally, faced with the need for sufficient information, which of the two special cases of the Bayesian filter, namely, the Kalman filter and the information filter, is the appropriate choice for resolving the ill-posedness of the sparse-coding problem? Without hesitation, the appropriate choice is the information filter, rooted in Fisher information. It is important to note, however, that the information filter does not create new information; rather, it facilitates the improved extraction of information contained in observables, hence, the improvement in sparse coding.

*The Divergence Problem:* Unfortunately, in their own respective ways, both the information filter and the Kalman filter suffer from the divergence problem, which could arise when the level of measurement noise power is relatively high. What happens in such a situation is the fact that the covariance matrix, or its inverse, violates the positive-definiteness requirement. To resolve the divergence problem, the recommended procedure is to use the Cholesky factorization to define the covariance matrix as the product of its square root and its transpose [27]. Moreover, the square root of the covariance matrix is propagated from one iteration to the next instead of the covariance matrix itself. Although by doing so, the square root of the covariance matrix may also violate the positive-definiteness requirement, the matrix product of the computed square root covariance matrix with its transpose recovers the desired positive-definiteness requirement of the covariance matrix, as it should be.

Regardless of whether the Cholesky factorization is needed, the information filter is well suited to stabilize the regularized sparse-coding problem by providing the needed sufficient information.

### D. Perceptual Attention for Improved Sparse Coding

Building on what has been said thus far, strong arguments can be made for the following statement:

“The performance of sparse coding is improved under the influence of perceptual attention through the use of information filtering.”

<sup>4</sup>The Kalman gain matrix is a recursion in the Kalman filtering algorithm [23].



This statement is justified on the following grounds in the paper by Amiri and Haykin [25].

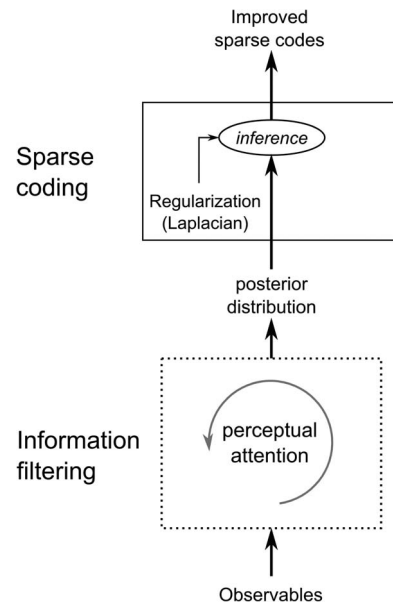
1)

“The ultimate objective of sparse coding is to resolve the source-separation problem, whereby relevant features extracted from the observables are favored, while at the same time the irrelevant features are rejected. Basically, the process, just described, is the very essence of decision-making, localized at the top layer of hierarchical perceptual memory.”

Recalling the principles of cognition discussed in Section I, decision making (under the principle of intelligence) builds on attention. It follows, therefore, that the perceptual memory must be endowed by attention to solve the source-separation problem. Here, we must recall the fact that intelligence is distributed throughout the entire cognitive dynamic system.

- 2) The second justification for supporting the influence of perceptual attention to improved sparse coding is on probabilistic grounds. To be specific, in the course of estimating a sparse representation, there is competition among different features contained in the noisy observables. It would be, therefore, logical to say that the higher the probability of occurrence of certain features in the observables, the greater will be the dynamic bias intended to select those particular features. On the other hand, the lower the probability of occurrence of other features in the observables, the more likely it is they will be ignored. Invoking the fact that dynamic bias in cognition is nothing but the basis of focused attention, we may, therefore, say that perceptual attention involves the ability to improve sparse coding.
- 3) Last, from the discussion presented in Section II, we recall that local perception–action cycles lead to attention. The information-filtering algorithm described in Section IV-B embodies a cyclic behavior within this algorithm, whereby the information vector is continually reinforced (i.e., improved in behavioral quality) in response to the incoming observables from one local cycle to the next. Here then is another plausible justification for the improved sparse coding under the influence of perceptual attention.

To sum up this discussion, through perceptual processing, the observables experience excitatory and inhibitory effects, which result in the selection of relevant features and the suppression of irrelevant features [25].



**Fig. 2. Perceptual cognit: Sparse coding augmented by information filtering.**

## E. Perceptual Cognit

Turning next to implementation of the improved sparse-coding algorithm with sufficient information provided for the sake of stability under varying environmental conditions, the information filter is inserted between the incoming observables and the traditional sparse coding, as depicted in Fig. 2. The bottom part of the figure shows the information-filtering block, and the top part of the figure depicts the essence of how sparse representation of the observables are inferred at the output regularized by a Laplacian distribution needed for sparseness. The end result of the whole process is an improved sparse-coding algorithm, as desired.

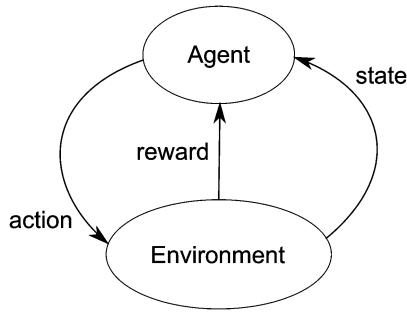
The entire scheme of Fig. 2, constitutes a perceptual cognit,<sup>5</sup> following the definition of a cognit described in Section II-B.

## V. DYNAMIC PROGRAMMING

Just as Section III was intended to set the stage for cognitive perception in Section IV, so it is with this section on Bellman’s dynamic programming [28], which provides the mathematical

<sup>5</sup>In [25], experimental results are presented to test the information processing power of perceptual attention, using real-life data recorded using the McMaster IPIX radar. Therein, it was demonstrated that the improved version of sparse coding under the influence of perceptual attention outperformed the traditional sparse-coding algorithm by learning better dictionary of features and generating better sparse codes.

Moreover, it is also of interest to note that radar data were complex valued, in that sample of the data was composed of amplitude and phase components. To have the design of a sparse-coding algorithm under the influence of perceptual attention, the so-called Wirtinger calculus was used to simplify the complex-valued computation.



**Fig. 3. Perception-action cycle, involving an agent (e.g., a robot).**

basis for cognitive control to be discussed in Section VI. Bellman's dynamic programming has been chosen here because it is a natural way of defining value functions from discounted rewards and then calculating optimal values; it is, therefore, widely used for optimal control in engineering applications. Most importantly, through a cognitive-inspired modification of dynamic programming, a new reinforcement learning algorithm emerges, which is described in this section; this new algorithm is endowed with some highly desirable properties, hence the practical relevance of the material presented in this section [29].

### A. State Decision Processes

Consider the block diagram of Fig. 3, as in the book by Sutton and Barto [30]. An agent (e.g., robot), responsible for decision making, operates in accordance with a Markov decision process that has the following characteristics:

- 1) in Markov decision processes, observations are commonly called the states of the environment, and the states are said to occupy a finite set;
- 2) for each state of the environment, there is a finite set of possible actions, out of which one particular action is selected by the agent;
- 3) every time the agent acts on the environment, a certain reward is given to the agent by the environment;
- 4) collectively speaking, states are observed, actions are taken, and rewards are delivered, all of which are performed at discrete times. Insofar as actions and states are concerned, Fig. 3 describes a "perception-action cycle."

Recognizing that the notion of state is of practical importance in the study of dynamic programming, the following definition is offered:

"The state of the environment is a summary of the entire past experience gained by the agent as a result of its continued interactions with the environment, such that the information necessary for the agent to predict future behavior of the environment is contained in that summary."

### B. Probabilistic Considerations

Let  $s_k$  denote the state of the environment,  $a_k$  denote the action taken by the agent on the environment, and  $r_k$  denote the reward provided by the environment in response to the action, where  $k$  denotes discrete time. To be realistic from a practical perspective, the environment is assumed to be stochastic (i.e., random). Hence, we may say that  $s_k$  is the sample value of some random variable.

With the environment being typically stochastic, transitions from one state to the next are naturally probabilistic. According to the crucially important Markov property discussed in Section III, the current state of the environment provides the necessary information for action to be taken by the agent on the environment. In other words, all future states and rewards are conditionally independent from past states.

Let  $a$  denote an action, which is also the sample value of a random variable of its own. Correspondingly, let  $\mathcal{P}_{ss'}^a$  denote the transition probability from the current state  $s = s_k$  to the new state  $s' = s_{k+1}$  due to the action  $a$  taken by the agent at time  $k$ . Then, involving the Markov property for state dynamics, the transition probability is formally defined by the conditional probability:

$$\mathcal{P}_{ss'}^a = \mathbb{P}[s'|s, a] \quad (4)$$

where  $\mathbb{P}$  is the operator for denoting probability. In accordance with probability theory, the transition probability satisfies the following two conditions:

$$\mathcal{P}_{ss'}^a > 0, \quad \text{for } s \text{ and } s', a \quad (5)$$

$$\sum_{s'} \mathcal{P}_{ss'}^a = 1, \quad \text{for all } s', a \quad (6)$$

where both  $s$  and  $s'$  lie in the state space. For a given number of states and given transition probabilities, the sequence of environmental states, resulting from action taken by the agent on the environment, forms a Markov chain in the course of time.

In the final analysis, the basic issue of interest in dynamic programming is summed up in the following statement:

"States of the environment are mapped into actions." This mapping is called the agent's policy and is denoted by  $\pi_k(s, a)$ , which is the probability that action  $a$  at time  $k$  occurs if, and only if, the state  $s_k = s$  at time  $k$ . The policy can be stationary or nonstationary; that is:

- if the policy is stationary, then the policy remains as  $\pi$  for all time  $k$ ;
- if, on the other hand, the policy is nonstationary, then the policy changes from one time step to the next, as in  $\pi_k$ .

### C. Bellman's Optimality Equation

In its most basic form, dynamic programming deals with finite-horizon problems, which means that the rewards are considered over a finite number of time steps. However, from the analytic perspectives, the preferred approach is to deal with infinite-horizon problems, where the rewards are considered over an infinite number of time steps. To simplify matters herein, we will merely highlight the essence of Bellman's optimality equation for infinite horizon problems in words, rather than equations. To this end, let  $Q(s, a)$  denote the action-value function for state  $s = s_k$  and action  $a = a_k$ , hence the statement:

"Function  $Q(s, a)$  testifies to the fact that it is 'good' as a result of interactions of the agent with its environment for a long time."

In [30], it is shown that the formula for  $Q(s, a)$  depends on the following points<sup>6</sup>:

- 1) expected immediate rewards
  - underlying dynamics of the interactions described in terms of probabilities;
  - the related  $Q$ -function in terms of  $s' = s_{k+1}$  and  $a'$ , that is,  $Q(s', a')$ ;
- 2) then, we have the following two summations in accordance with (7):
  - the first summation is over all possible actions;
  - the second summation is over all possible states.

The optimal  $Q$ -function, denoted by  $Q^*(s, a)$ , is obtained by maximizing the sum of all terms with respect to action  $a$ . Unfortunately, the end result of this maximization is an exponential growth in computational complexity, which is referred to as the curse of dimensionality [31].

What is truly remarkable is the following: The idea of dynamic programming was pioneered by Bellman in a book published under that same title in 1957; yet, almost 60 years later, that very idea is still occupying the attention of engineers and scientists on account of its mathematical foundation, particularly when optimal control is the issue of interest. In Section VI, we describe a new reinforcement learning algorithm that is derived by modifying Bellman's

<sup>6</sup>In [30], Bellman's optimality equation is defined for the state-value function on page 70. The corresponding version of this equation for the action-value function (i.e., the  $Q$ -function) is left as an exercise on page 72; the equation for the  $Q$ -function, so derived, is described as follows:

$$Q(s, a) = \sum_{s'} \mathcal{P}_{ss'}^a \left[ \mathcal{R}_{ss'}^a + \gamma \sum_{a'} \pi_k(s', a, a') Q(s', a') \right] \quad (7)$$

where  $\mathcal{R}_{ss'}^a$  is the expected reward for the transition from state  $s$  to state  $s'$  caused by action  $a$ . In the procedure described in [29], the modified version of Bellman's equation is correspondingly defined as follows:

$$Q(a) = r_{k+1} + \gamma \sum_{a'} \pi_k(a, a') Q(a') \quad (8)$$

where  $\gamma$  is a discount factor, and  $\mathcal{P}_{ss'}^a$  is (4). Linearity of the modified version of Bellman's equation, with respect to action  $a$ , follows directly from (8).

optimality equation without any approximation whatsoever, yet the algorithm retains optimality.

## VI. COGNITIVE CONTROL

In Section I, we described the principles of human cognition, namely, perception, memory, attention, and intelligence, which adequately covered the material presented on cognitive perception in Section IV. Continuing on, Section V on dynamic programming, culminating in the celebrated Bellman's optimality equation, prepares us for this section on cognitive control. In a loose sense, we may, therefore, view Section V as a "transition point" in mathematical terms from cognitive perception to cognitive control.

To reinforce this transition in a complimentary manner, we propose to look to some relevant aspects of the prefrontal cortex, which, in one form or another, influences the discussion presented not just in this section, but the follow-up two sections on the probabilistic reasoning machine and risk control.

### A. The Prefrontal Cortex

From the point of view of contemporary neuroscience, cognitive control is clearly within the physiological purview of the prefrontal cortex [32], [33]. The principal functions of cognitive control are the following: working memory, attentional set, error monitoring, and decision making. All four of them have a future perspective, as they serve the overarching role of that cortex, which is the organization of new and complex goal-directed actions, hence serving the preadaptation function, on which more will be said in Section VIII. Of course, the prefrontal cortex cannot perform any of these top-down functions without other cortical and subcortical structures, but none of those functions can be performed correctly and efficiently without the functional integrity of the prefrontal cortex at the summit of the global perception-action cycle. The execution of any given sequence of behavior takes place by a processing cascade down the executive hierarchy of cortical areas and the executive cognits they harbor individually [34]–[36].

We summarize brief descriptions of the principal functions of prefrontal cortex [33].

- 1) Working memory. By definition, this principal function is a prospective function. In effect, working memory is the temporary retention of critical information for the pursuit of a goal or the solution of a problem [37]. More specifically, it serves as the temporal integrator—to bridge any discontinuities—at the top of the global perception-action cycle. Reentry and feedback between areas, at the highest cortical level, are essential to the good functioning of working memory. Reverberation in reentrant circuits is

probably at the foundation of the neurophysiology of working memory [33].

- 2) Attentional set. This is the priming of sensory and motor control for the action in preparation and its expected adaptive consequences. Attentional set, consisting of perceptual attention and executive attention, is guided by feedback from previous actions. It includes the two reciprocal aspects of attention:
  - a) the focus on present cognitive demands (with limited neural resources);
  - b) the exclusionary aspect of inhibition of irrelevant (distracting) information.
- 3) Error monitoring. Also by reentrant feedback from executive structures and the result of every action in a sequence, the medial prefrontal cortex will receive the relevant information to correct prediction errors and to guide subsequent actions to their goal. There is evidence that the same prefrontal structure (anterior cingulate cortex) involved in monitoring errors is in charge of their prevention [38], [39].
- 4) Decision making. Finally, decision making is the choice of intended action, by the prefrontal cortex, often in the presence of uncertainty or ambiguity. The prefrontal cortex decides on the basis of several factors:
  - a) influences that arise from the current environment, from the internal milieu, and limbic system (affect and motivation);
  - b) from cortical cognits of past memory, be it perceptual or executive.

The decision may be simple and straightforward, or else the result of complex computation of winner-takes-all. In any case, it has to be preadaptive, after probabilistic appraisal of risks, benefits, and past history. More will be said on these four functions in Sections VII and VIII.

## B. The Imperfect State Information Problems

For now, with cognitive control in mind, we refer back to Bellman's dynamic programming discussed in Section V, the applicability of which rests on the following requirement:

"The agent (e.g., robot) has access to exact values of the state of the environment at all times."

This statement applies equally well to reinforcement learning. However, it would be unrealistic in practice to expect that every controller in the world, acting on its environment, would satisfy this highly stringent requirement. We say so, because it is feasible for some state variables of the environment to be inaccessible to the cognitive controller; examples of such situations, to name just a few, include:

- the human brain;

- cognitive radar;
- cognitive sonar;
- the power grid.

In volume I of his classic book entitled *Dynamic Programming and Optimal Control*, Bertsekas [40] referred to problems, where the controller does not have access to some or all of the state variables of the environment, as "imperfect state information problems."<sup>7</sup> To overcome this practical difficulty, in a clever way, Bertsekas [40] developed appropriate "transformations," whereby new dynamic programming algorithms were formulated for solving imperfect state information problems. Unfortunately, those new algorithms were computationally far more demanding than their counterparts that involve perfect state information; we thus have another example of the no-free lunch theorem. Summarizing this discussion:

"Unless a dynamic programming or reinforcement learning algorithm has access to exact values of all the state variables of the environment, the algorithm is confronted with the imperfect state information problem."

## C. Entropic State of the Perceptor

To find a solution for the imperfect state information problem in a cognitive dynamic system, we look to the perception errors produced in a particular layer of the system. As the information content of the perception errors is reduced, the accuracy of the perceptor in the pertinent layer is correspondingly improved. In other words, the information content of the perception errors provides a measure of how perfect the state of the perceptor is. Intuitively, it is appealing that we look to Shannon's entropy<sup>8</sup> [43], [44] as the measure for

<sup>7</sup>An illustrative example addressing the imperfect state information problem is presented in [41]. Therein, a tracking radar with perception-action cycle only was used to track a falling object in space. Confronted with the imperfect state information problem, and using the transformation described in [40], computational complexity of the resulting algorithm was so high that the study was limited to the use of "dynamic optimization" with practically "no predictive capability."

<sup>8</sup>Consider a continuous random variable  $X$ , with  $x$  denoting a sample value of  $X$ . According to Shannon's information theory, the entropy of  $X$  is defined by

$$H(X) = \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx$$

where  $p(x)$  is the probability density function of the random variable  $X$ , and  $\log$  is the symbol for logarithm. For a discrete random variable, the entropy is defined by

$$H(X) = \sum_{k=-\infty}^{\infty} p(x_k) \log \frac{1}{p(x_k)}$$

where  $x_k$  is the sample value of  $X$ , defined at discrete time  $k$ .

The idea of perception entropic state was first described in [16]. Most importantly, the first demonstration confirming the role of perception entropic state as the legitimate input state for the cognitive controller was presented in [42].

quantifying the information content of the perception errors. We may thus make the following statement:

“For a layer in a cognitive dynamic system, an entropic state of the perceptor in that layer provides the desired measure of the perception errors.”

Furthermore, the perception entropic state provides the feedback information from the perceptor to the controller, thereby linking them together in each layer of the cognitive dynamic system.

Thus, just as the cognitive perceptor in a layer of the system has direct access to the environmental observables, the cognitive controller in that layer has direct access to the perception entropic state. Accordingly, the imperfect state information in a cognitive dynamic system is no longer a problem.

#### D. Multiple-State Model of Cognitive Dynamic Systems

Moreover, the cognitive controller is also the subject to unavoidable errors of its own. To be more specific, using arguments similar to those used for the perception entropic state, we may go on to make the dual statement:

“The entropic state, pertaining to the control errors in a particular layer of the cognitive dynamic system, is a state of the cognitive controller.”

The role of perception entropic state (i.e., feedback information) is to provide the input to the cognitive controller for action on the environment. In light of the entropic material presented thus far, we may now postulate the multiple-state model of cognitive dynamic systems, composed of three parts, as follows:

- 1) state-space model of the environment, which consists of two equations as described in Section III;
- 2) vector of perception entropic states, each element of which pertains to a perceptor in a particular layer of the hierarchical cognitive dynamic system;
- 3) vector of control entropic states, each element of which pertains to a controller in the corresponding layer of the hierarchical cognitive dynamic system.

Insofar, as the cognitive dynamic system is concerned, the vectors of perception and control entropic states may well be distributed over their respective manifolds in a multidimensional entropic state space.

#### E. Revisiting Bellman’s Optimality Equation

Continuing the discussion presented in Section VI-C, the stage is set for us to revisit Bellman’s optimality in Section V, and build on the following two ideas:

- 1) the entropic state of the perceptor;
- 2) the fact that reward is a function of two states, namely, the current entropic state and its immediate past value.

Let  $r_k$  denote the reward at time  $k$ ; and let  $s_k$  and  $s_{k-1}$  denote the perception entropic state at time  $k$  and  $k-1$ , respectively. We may then introduce the following relationship between the reward and states:

$$r_k = g(s_k, s_{k-1}) \quad (9)$$

where  $g$  is an invertible function (e.g., logarithmic function). Hence, we may make the following statement [29]:

“Knowing only the initial state  $s_0$  and the first reward  $r_1$ , a cognitive version of Bellman’s optimality equation can be derived from its classic regular form, as shown in (8).”

Accordingly, following the underlying mathematics presented in [29], relevant points of which were summarized in footnote 6, we end up with the derivation of a new reinforcement learning algorithm, which has some highly desirable properties.<sup>9</sup>

- *Property 1.* Computational complexity of the new reinforcement learning algorithm follows a linear law, which is the best that it could be; theoretical verification of this property follows directly from (8).
- *Property 2.* Unlike traditional reinforcement learning algorithms, there is no approximation whatsoever in deriving the new reinforcement learning algorithm.
- *Property 3.* With the imperfect state information problem no longer being relevant, the new reinforcement learning algorithm is no longer restricted to having access to all the state variables of the environment.
- *Property 4.* Recognizing that the classical Bellman dynamic programming algorithm is convergent to an optimal policy [28], and the new reinforcement learning algorithm is a special case of it, it follows that the new reinforcement learning algorithm is also convergent.

The properties of the new reinforcement algorithm, just described, are profound from a practical perspective, prompting us to make the statement:

“Since the new reinforcement learning algorithm has been developed within a cognitive dynamic system, and in cognitive neuroscience we do speak of cognitive functions, henceforth, we refer to the new algorithm as the cognitive reinforcement learning algorithm.”<sup>10</sup>

<sup>9</sup>In [29], two different illustrative examples are presented: a stochastic network and a cognitive radar; both examples support the statements described under Properties 1–4.

<sup>10</sup>The information processing capacity of the cognitive reinforcement learning algorithm has been demonstrated using Monte Carlo simulations for two entirely different experiments [29]: 1) observability of a stochastic network, where only a small number of network’s nodes were accessible to the cognitive controller; and 2) demonstration of the superior performance of the cognitive reinforcement learning algorithm over traditional reinforcement learning algorithm, namely, Q-learning [30], with the experiment being performed on a cognitive radar.



## F. Practical Utilities of the Cognitive Controller

**Rewards:** Reward is an essential element in the formulation of reinforcement learning. In light of (8), the reward for the cognitive reinforcement learning algorithm is defined as follows:

$$r_k = g_k(H_k; \Delta_1 H) \quad (10)$$

where  $H_k$  is the perception entropic state of the cognitive perception at time  $k$ , and  $\Delta_1 H$  is the incremental difference, defined by

$$\Delta_1 H = H_{k-1} - H_k \quad (11)$$

where it is noted that, usually, the perception entropic state decreases with time  $k$ . But, it is feasible for the entropic reward to be negative, which means punishment. Note also that the use of subscript  $k$  in the invertible function  $g_k$  simply means that this function is updated from one global perception–action cycle to the next. Moreover, the use of planning plays a key role in defining the control policy (i.e., mapping steps into actions) from one global perception–action cycle to the next. Naturally, planning requires the formulation of a probabilistic model of the controller and/or the environment, which is needed for the prediction of future rewards.

**Explore–Exploit Tradeoff:** When planning is the primary issue of interest, then we have a pure exploration strategy in the following sense:

“The cognitive controller visits every action in the action space, so that the controller makes better action selections in the future.”

In direct contrast, when optimality is the primary issue of interest, then we have a pure exploitation strategy in the alternative sense:

“In this case, the cognitive controller focuses attention only on those actions in the action space, which are likely to be the actions of special interest to the controller.”

Clearly, these two pure strategies are in conflict with each other, hence the terminology “explore–exploit tradeoff.”

As a compromise between the two pure strategies, a commonly recommended procedure in practice is a mixed strategy, which involves both exploration and exploitation in the following twofold sense:

- 1) an explore rate, denoted by  $\epsilon$ , is specified ahead of time, where  $\epsilon$  is the fraction of global perception–

action cycles for which the cognitive controller makes decisions at random, meaning exploration<sup>11</sup>;

- 2) for the remainder of global perception–action cycles, the cognitive controller works on exploitation.

The intuitive appeal of this mixed strategy is that the cognitive controller maintains a certain degree of forced exploration, while during the global perception–action cycles allocated for exploitation, the cognitive controller focuses attention on those actions that are likely to be of most practical value. This heuristic strategy is called the  $\epsilon$ -greedy strategy [45].

From the discussion just presented, we see that the explore–exploit strategy could be a facilitator of control attention.<sup>12</sup>

## G. Management of Resources

The management of resources is naturally application dependent. Nevertheless, the one resource that is common to all practical applications of cognitive dynamic systems is computational complexity. For the most efficient use of computational resources, the optimal solution is to adopt algorithms that follow a linear law. With this important point in mind, we may summarize the practical utility of the main algorithms discussed thus far, as follows:

- 1) for cognitive perception discussed in Section IV, we say that the algorithm we focused on was the improved sparse-coding algorithm that operated under the influence of perceptual attention;
- 2) as for cognitive control discussed in this section, we say that the algorithm that featured most prominently in the discussion is the cognitive reinforcement learning algorithm.

Although, indeed, these two algorithms address perception and control that are the dual of each other, they both share a common characteristic: linear law for computation.

## H. Cognitive Reinforcement Learning From Engineering and Neuroscience Perspectives

Referring back to the four properties of cognitive reinforcement learning described in Section VI-E, we may readily say that this new reinforcement learning algorithm is superior to traditional reinforcement learning algorithms [30].

However, when we examine the new reinforcement learning algorithm from the neuroscience perspective, we

<sup>11</sup>The problem with exploration is the fact that the random selection of an action may lead to a poor decision. We may mitigate this problem by using the Boltzmann exploration, where an action is selected with a probability proportional to the estimated value of an action [45]. This alternative approach to exploration derives its name from the Boltzmann distribution, well known in thermodynamics.

<sup>12</sup>In [46], a rather simple exploit–explore strategy was used for control in a cognitive radar with memory, and dynamic optimization was used for the controller. Therein, it was demonstrated that the explore–exploit strategy can improve behavior of the transmitted waveform significantly. Unlike the restricted use of explore strategy by itself, the explore–exploit strategy resulted in a smooth transition in the transmitted waveform from one global perception–action cycle to the next, at the price of small degradation in the learning rate.



have to be aware of precautionary realities [47], summarized as follows:

- 1) after almost 100 years of intense experimental and theoretical investigations, at the level of neurons and synapses, we are not that much further ahead;
- 2) the failure of traditional reinforcement learning stems from the fact that learning behavior, at the level of neurons and synapses, is rather complex, which should not be surprising.

It is, therefore, difficult to know where the new reinforcement learning algorithm fits into the scheme of things, viewed from a neuroscience perspective.

## VII. PROBABILISTIC REASONING MACHINE

In this section, we discuss the probabilistic reasoning machine, which, in a reciprocal manner, couples cognitive perception and cognitive control together, as depicted in Fig. 1. The machine is so called, because its function within the cognitive dynamic system is based on probabilistic reasoning, a branch of probability theory.

In operational terms, particularly in a typically stochastic (i.e., random) environment that is nearly always nonstationary, the probabilistic reasoning machine mediates between the cognitive controller and the cognitive perceptor, realizing the following objective:

“When the cognitive dynamic system operates in a nonstationary stochastic environment, self-organized and stable behavior are assured on a continuing time basis.”

With the environment being stochastic, it is compelling that we bring probability theory into play. As for the issue of nonstationarity, the way to tackle it is for the mediating machine to exploit probabilistic reasoning, which is how the rest of this section will proceed.

Moreover, recognizing that both the cognitive perceptor and the cognitive controller make errors, the discussion will, therefore, focus on the roles of perception entropic state and control entropic state as the respective measures for those errors. Indeed, when we speak of overall system stability, this pair of entropic states feature prominently in the discussion. Mentioning entropic states of the probabilistic errors, perception as well as control, reminds us of the error monitoring function, which was considered to be one of the principal functions in the prefrontal cortex discussed in Section VI. Not only that, the remaining three functions, attentional set, working memory, and decision making, also play key roles of their own in the operation of the probabilistic reasoning machine. It may, therefore, be said that from an engineering perspective, the probabilistic reasoning machine is inspired by the prefrontal cortex.

### A. Probabilistic Reasoning

As mentioned previously, probabilistic reasoning is a branch of probability theory, the purpose of which is twofold [48], [49]:

- 1) tackling the unavoidable presence of external uncertainties in a satisfactory manner;
- 2) providing the cognitive dynamic system an information processing capacity of deductive logic.

Consider a sample layer of the cognitive dynamic system, depicted in Fig. 4. Three respective units of the system feature in the figure. The components pertaining to the controller and perceptor are reciprocally coupled via the corresponding unit of the probability reasoning machine.

Before proceeding forth, it is instructive that we refer back to the internally composite perception–action cycle, discussed in Section I. Most importantly, this cycle embodies the cognitive perception, probabilistic reasoning, and cognitive control, eschewing the environment. In effect, this particular cycle is indeed responsible for the attentional set, made up of both perceptual attention and control attention, hence its practical importance in the material to follow.

In an overall sense, the primary function of the probabilistic reasoning machine is highly profound on account of two essential requirements, recalling what was said earlier on in this section:

- 1) overall stability of the cognitive dynamic system is maintained on a layer-by-layer basis;
- 2) self-organized behavior of the system is assured throughout the system.

To realize these two functions continuously, we look to deductive logic, which operates on the basis of the IF–THEN rule.

### B. Decision Making for System Stability

To justify how stability is maintained at the layer level, let  $H_k^{(p)}$  denote the entropic state of the perceptor at time  $k$ , and let  $H_k^{(c)}$  denote the corresponding entropic state of the controller. For overall stability of the cognitive dynamic system, the following pair of constraints would have to be imposed on these two entropic states, as shown by

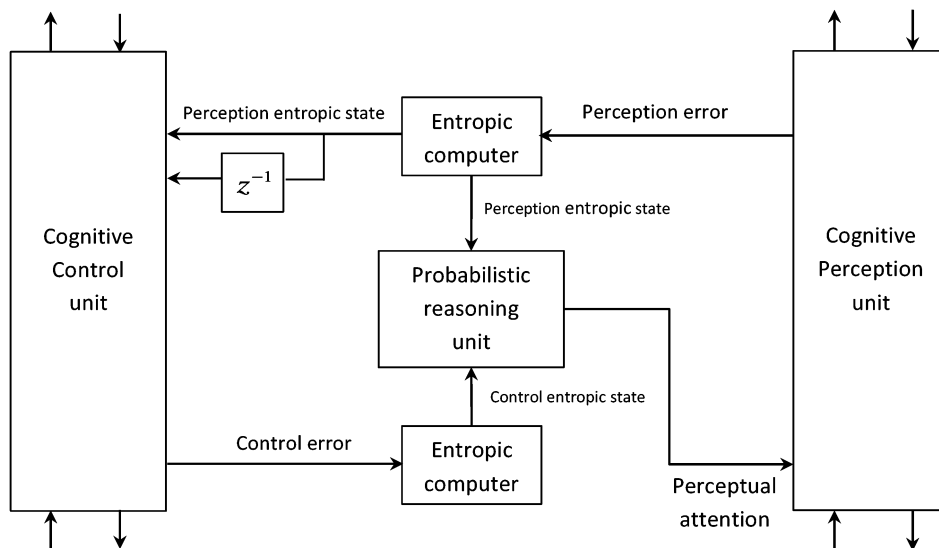
$$\text{a) } 0 < H_k^{(p)} \leq \epsilon_p \quad (12)$$

$$\text{b) } 0 < H_k^{(c)} \leq \epsilon_c \quad (13)$$

where  $\epsilon_p$  and  $\epsilon_c$  are prescribed upper bounds. This pair of constraints would have to be satisfied on a layer-by-layer basis, if we are to realize the two primary functions described previously.

It is important to note that neither one of the two entropic states would ever assume the value zero, because of two reasons:

- 1) there would then be no feedback information from the perceptor to the controller;



**Fig. 4. Functional block diagram of probabilistic reasoning unit for a layer in the cognitive dynamic system; the block  $z^{-1}$  represents time-unit delay.**

- 2) if point 1) is true, then there would also be no perceptual attention applied to the perceptor by the controller.

However, neither of these two conditions would ever arise in practice because of the persistent presence of perception errors and control errors, no matter how small.

In any event, there are four different probabilistic scenarios that would have to be considered:

- *Scenario 1 (Perfect Stability)*: The two constraints a) and b) on the perception and control entropic states are both satisfied, and there is, therefore, no more to be said.
- *Scenario 2*: Constraint a) pertaining to the perceptor is violated, that is

$$\begin{aligned} \text{a)} \quad & 0 < \epsilon_p < H_k^{(p)} \\ \text{b)} \quad & 0 < H_k^{(c)} \leq \epsilon_c. \end{aligned}$$

To stabilize the system under this second scenario, the perceptual attention in the attentional set comes to the rescue:

“If the second probabilistic scenario occurs, then the perceptual attention tends to bring the perception errors downward, from one global perception–action cycle to the next.”

Accordingly, the original condition a) of scenario 1 is reestablished within a limited number of global perception–action cycles.

- *Scenario 3*: Constraint b) pertaining to the controller is violated, that is

$$\begin{aligned} \text{a)} \quad & 0 < H_k^{(p)} \leq \epsilon_p \\ \text{b)} \quad & 0 < \epsilon_c < H_k^{(c)}. \end{aligned}$$

To stabilize the system under this third scenario, control attention in the attentional set comes to the rescue:

“If the third probabilistic scenario occurs, then the control attention tends to bring the control errors downward, from one global perception–action cycle to the next.”

Here again, after a limited number of global perception–action cycles, the original condition b) of scenario 1 is reestablished.

- *Scenario 4*: Both constraints are violated, that is

$$\begin{aligned} \text{a)} \quad & 0 < \epsilon_p < H_k^{(p)} \\ \text{b)} \quad & 0 < \epsilon_c < H_k^{(c)}. \end{aligned}$$

Under this extreme condition, the attentional set comes to the rescue:

“If the somewhat less probabilistic scenario 4 occurs, then the attentional set brings both perception errors and control errors

downward simultaneously from one global perception–action cycle to the next.”

Finally, the two original conditions a) and b) of scenario 1 are reestablished, again after a somewhat larger number of global perception–action cycles.

The conclusion to be drawn from this deductive logic under probabilistic reasoning is that the constraints under (12) and (13) will be maintained, in one form or another. In other words, the overall system stability as well as its self-organized behavior are assured on a continuing basis, except for the odd short periods where one of the three scenarios 2, 3, or 4 occurs. It is, therefore, on account of this profound information processing capability, that we may refer to the probabilistic reasoning machine to be at the heart of the cognitive dynamic system.

### C. Summarizing Remarks

In Section VI, we said that the prefrontal cortex is characterized by four principal functions: working memory, attentional set, error monitoring, and decision making. Indeed, all these four principal functions are basic to a probabilistic reasoning machine.

- 1) When we speak of working memory, we typically think of short-term memory (i.e., temporary retention of critical information). Indeed, examination of Fig. 4 reveals the existence of short-term memory, exemplified by the pair of perception entropic states in the left-hand side of Fig. 4, namely, the current perception entropic state and its immediate past value due to operator  $z^{-1}$ .
- 2) As for the attentional set, it is brought into play by the internally composite perception–action cycle that embodies all the three functional blocks in Fig. 4. Note also that this figure includes a local perception–action cycle that results in perceptual attention; this additional local cycle enhances the attentional capability of the cognitive perceptor, which, in reality, is needed to further improve the separation of relevant information from irrelevant information.
- 3) Error monitoring is taken care of, first, by the use of Shannon’s entropy for measurements of perception and control errors, which are expressed in terms of entropic states. Second, this transformation makes it relatively straightforward for the probabilistic reasoning machine to monitor the perception and control errors on a continuing-time basis.
- 4) Finally, the application of the IF–THEN rule to the last three scenarios addressed in Section VII-B is the essence of decision making, which manifests itself in overall system stability and self-organization that are the hallmarks of cognitive dynamic systems.

## VIII. RISK CONTROL

In this section, we discuss a challenging problem that is the most difficult of them all, namely, risk control. The solution to this problem is inspired by the prefrontal cortex in the brain. To be more precise, the prefrontal cortex involves a proactive function, defined simply as follows:

“Prediction is followed by preadaptation.”

Prediction is performed under planning in the motor control. As for preadaptation, it is rooted in the ability of the prefrontal cortex to temporally organize a “new” adaptive information processing mechanism, which is intended for overseeing goal-directed behavior in the future. The issue of prediction is relatively straightforward to understand. On the other hand, the issue of preadaptation is where the challenge is. With this brief background on the prefrontal cortex, we are ready to address the following challenge from an engineering perspective<sup>13</sup>:

“How do we construct an adaptive information processing mechanism within the motor control, in such a way that it mimics the predict and preadapt functions in the prefrontal cortex, with decision making as the final product for action on the environment.”

Recall that decision making is the last one of the principal functions in the prefrontal cortex, discussed in Section VI.

### A. Risk Control: Definition

We begin the discussion by identifying two learning curves for a cognitive dynamic system:

- 1) the perception learning curve plots the entropic state of the cognitive perceptor versus time;
- 2) similarly, the control learning curve plots the entropic state of the cognitive controller versus time.

Consider then a cognitive dynamic system that operates in a nonstationary stochastic environment; the system is then said to be operating in regular fashion if, and only if, both learning curves remain above their respective prescribed thresholds across time; quite likely, these two thresholds could be different in their respective layers.

Suppose, next, a disturbance occurs in the environment unexpectedly, strong enough to push one, the other, or both, learning curves below the prescribed thresholds. Furthermore, suppose the disturbance lasts long enough to have a serious impact on the behavior of the cognitive controller.

<sup>13</sup>The differentiation of “pre” in preadaptation and prefrontal should be carefully noted: in the former, it is temporal; whereas in the latter, it is spatial. In any event, the work done on preadaptation from an engineering perspective should be beneficial in the study of preadaptation in neuroscience.

Under these conditions, we may now define the function of risk control to be that of satisfying the following pair of requirements:

- 1) the preadaptive control function built into the motor control comes into play immediately, with the objective of maintaining the goal-directed behavior of a target in the next global perception–action cycle, on a cycle-by-cycle basis;
- 2) once the disturbance is finished, both learning curves reestablish their regular trajectories in a smooth manner.

The challenge here is how to mimic the “predict and preadapt” function in the prefrontal cortex, so as to control the risk attributed to the disturbance.<sup>14</sup>

### B. Preadaptive Control Mechanism

A block diagrammatic description of motor control, the last functional block in the cognitive control, is depicted in Fig. 5. Under the influence of the input from the bottom of the cognitive control hierarchy, the policy searches the exploit–explore library for “possible” actions for use in the future (i.e., prediction).

The remaining part of the figure is not only new, but also it is aimed at mimicking the preadaptation function in the prefrontal cortex. Specifically, the preadaptive control mechanism culminates with decision making, in accordance with the prefrontal cortex. This mechanism consists of the following components.

- 1) The functional block to the right of the figure, the function of which is to store a sequence of experiences gained in the past; the window is adjustable, depending on how far back we go to the past.
- 2) Tapped-delay line, which consists of  $N + 1$  time-unit delay elements, each denoted by the symbol  $z^{-1}$ ; the first  $z^{-1}$  delays the optimal action to be taken on the environment by one time unit. The remaining  $N$  time units account for storing experiences gained from past interactions of motor control with the environment over the course of  $N$  perception–action cycles, including the current one.

<sup>14</sup>In a rather simple, yet informative, experiment described in [15], three radar configurations were tested for the impact of a disturbance on their respective learning curves. Specifically, a filtered white Gaussian noise of limited duration was added to the process equation of the state–space model, over and above the process noise. The disturbance lasted for about 2 s within 6 s of the experiment, which involved the following. 1) The simplest radar used the Kalman filter in the receiver; there was no perception–action cycle. In this part of the experiment, the learning curve changed the course of its trajectory by moving upward and was unable to recover. 2) In the second part of the experiment, perception–action cycle was added to the radar. Here again the learning curve changed its trajectory, and when the disturbance finished, the learning curve seemed to move toward its original trajectory, but apparently at a very slow rate. 3) In the third part of the experiment, one layer of memory was added to the radar configuration used in part 2). In this case, the learning curve deviated slightly from the course of its regular trajectory, for the entire duration of the disturbance; however, as soon as the disturbance was finished, the radar recovered its original trajectory, thanks to the use of memory for predicting the consequences of actions.

- 3) MAP unit, the purpose of which is to select the best experience gained within the past (anywhere within the window), by processing the complete sequence of  $N$  past experiences.
- 4) Nearest neighbor classifier [50], which is driven from the left by the predicted set of selected actions put forward by the motor control policy, and driven from the right by output of the MAP unit. Finally, the nearest neighbor classifier picks that particular action in the set selected by the policy, which is the closest to the output of the MAP unit.<sup>15</sup>
- 5) The closest action picked from the library provides the action for the next global perception–action cycle, hence the notion of preadaptation.

Important highlights of the preadaptive control mechanism are as follows.

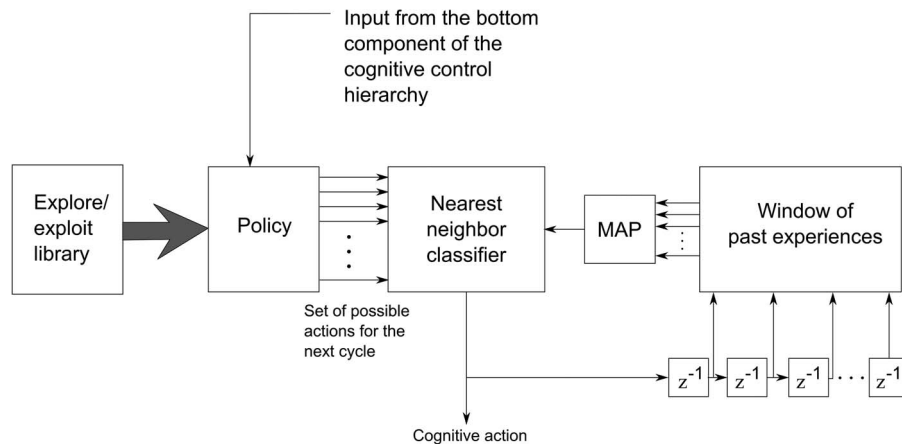
- 1) Experiences gained in the past over a sequence of global perception–action cycles play a key role in the decision-making process.
- 2) The combination of policy, tapped-line store of experiences gained in the past, MAP unit, and nearest neighbor classifier, they all constitute a closed-loop feedback structure for decision making; the end result is not just the choice of the next action, but also control (executive) attention.
- 3) There are essentially two optimal computations being performed within the feedback loop: one aimed at selecting the best past experience, and the other aimed at selecting the optimal cognitive action for the next global perception–action cycle, which is a rarity in engineering applications.

### C. Risk Brought Under Control

We are now ready to describe how risk, arising from the unexpected occurrence of a disturbance in the environment, is actually brought under control by proceeding as follows.

- 1) Once a disturbance occurs in the environment, there will be a corresponding change not only in the observables, but also in the perception entropic state.
- 2) Change in the perception entropic state induces a corresponding change of its own in the entire operation of the cognitive controller, thereby bringing the preadaptive control mechanisms in the motor control into play.
- 3) At this point, the preadaptive control mechanism, empowered by control (executive) attention, begins to maintain the goal-directed behavior of the target of interest from one global perception–action cycle to the next.

<sup>15</sup>It is noteworthy that the explore–exploit tradeoff used for cognitive radar in [46] is a special case of the complex preadaptation control mechanism in Fig. 5.



**Fig. 5. Functional block diagram of motor control, augmented with the preadaptive control mechanism.**

- 4) Finally, once the disturbance comes to an end, the perturbed target trajectory joins up with the regular trajectory in a smooth manner.

In other words, insofar as the trajectory is concerned, the perturbed target trajectory deviates marginally from its regular behavior, thanks to the preadaptive control mechanism. In this context, we may say that in light of footnote 14, the information processing power of the preadaptive control mechanism is reinforced by the ability of the perceptual memory to predict consequences of action taken by the motor control.

#### D. Illustrative Figure Highlighting the Past, the Present, and the Future

Fig. 6 presents another way in which the preadaptive control mechanism performs its decision making: On the left-hand side of the figure, we have a set of possible actions selected from a large exploit–explore library, under the influence of motor-control policy. On the right-hand side of the figure, we have the best experience selected from a set of past experiences by the nearest neighbor classifier.<sup>16</sup>

#### E. System Stability and Risk Control

These two basic functions of a cognitive dynamic system distinguish themselves vividly from each other, as illustrated in Fig. 7.

- 1) System stability discussed in Section VII on the probabilistic reasoning machine is governed by the internally composite perception–action cycle<sup>17</sup> as well as local perception–action cycles, none of

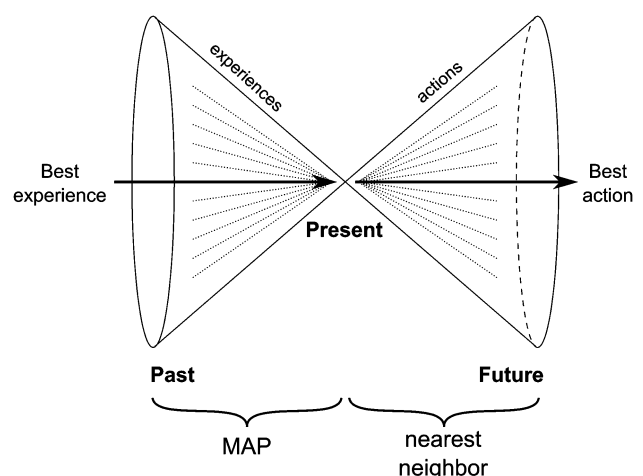
which involves the environment, as depicted in the left-hand side of the figure.

- 2) On the other hand, risk control is governed by the global perception–action cycle and the preadaptation cycle, both of which do involve the environment, as depicted in the right-hand side of the figure.

In effect, Fig. 7 illustrates that the two important functions of a cognitive dynamic system are entirely different: the first one is of an internal kind, and the second one is of an external kind. Furthermore, the pairs of cycles represent system stability and risk control operating in opposite directions, hence the separation of system stability and risk control.

Two other noteworthy points are:

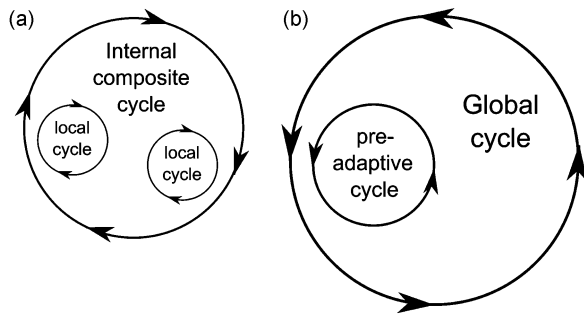
- 1) risk control could fail, while system stability is maintained;



**Fig. 6. Display of past experiences on the left and, future actions for selecting the best action.**

<sup>16</sup>The diagram in Fig. 6, where one side converges from the past and the other side diverges to the future, mimics a similar diagram of the brain in [51], origin of which is attributed to the Philosopher Karl Popper.

<sup>17</sup>In the literature on cognitive neuroscience, this cycle is described as an internal perception–action cycle, which runs from the motor system to the sensory system. The most current, theoretical expression of the internal shunt is due to Clark [52], where it is referred to as an “action-oriented” predictive processing system.



**Fig. 7. Graphical display of the four distinctive cycles of a cognitive dynamic system. (a) System stability. (b) Risk control.**

- 2) if, however, the cognitive dynamic system fails due to a malfunction within the system and thereafter becomes unstable, risk control would fail with it.

## IX. CONCLUSION

The cognitive dynamic system represents a new information processing model of the environment (world), which is inspired by the human brain. In the final section of this paper, we list ten major inferences from using the model, each of which has two objectives in mind and each one in its own way:

- 1) the potential for engineering applications;
- 2) the corresponding neuroscience inference, on which the model may be anchored.

The ten major inferences are as follows.

- 1) *Bayesian paradigm*: When the issue of interest is perception of the environment, the Bayesian paradigm finds the right home in neuroscience as well as engineering.
- 2) *Sparse coding*: Our model predicts sparse coding with increased perceptual attention. A comparable effect has been observed in neurons of sensory associative cortex.
- 3) *Perception entropic state*: The entropic state, based on perception errors, enables the controller of a cognitive dynamic system to operate effectively and efficiently. We infer that a similar state should exist in the brain.
- 4) *New reinforcement learning algorithm*: The new reinforcement-learning algorithm, a special case

of Bellman's optimality equation, is a game changer for engineering applications. At this stage, however, it is too early to know how it fits in the context of learning behavior in cognitive neuroscience.

- 5) *Probabilistic reasoning machine*: This new machine mediates the transactions between cognitive perception and cognitive control, much as it occurs in the brain at the top of the internal perception-action cycle.
- 6) *Internal perception-action cycle*: In a cognitive dynamic system, the internally composite perception-action cycle, eschewing the environment, plays a key role in the probabilistic reasoning machine, just as the internal shunt in the perception-action cycle that goes from the motor to the sensory system.
- 7) *Risk control*: The control of risk, by far the most challenging of them all, is accounted for in motor control by a novel preadaptive control mechanism in the form of a closed-loop feedback structure, which mimics prediction followed by preadaptation in the prefrontal cortex.
- 8) *System stability and risk control*: The primary functions of a cognitive dynamic system, namely, system stability and risk control, are different, just as they are in the brain.
- 9) *Network of cognits*: The cognitive dynamic system is a complex network made up of perceptual, probabilistic reasoning, and control cognits, just as it is in the brain.
- 10) *Modeling complex physical problems*: The cognitive dynamic system could be viewed as an information processing computer capable of solving complex physical problems characterized by nonlinearity, nonstationarity, and non-Gaussianity in an online manner, mimicking the brain in this powerful capability. ■

## Acknowledgment

The authors would like to thank Prof. K. Friston (University College London, London, U.K.) and Prof. T. J. Sejnowski (Salk Institute, La Jolla, CA, USA) for their helpful comments on earlier versions of the paper. They would also like to thank the Ph.D. students A. Amiri and M. Fatemi (McMaster University, Hamilton, ON, Canada) for their helpful comments on many versions of the paper.

## REFERENCES

- [1] P. L. Broca, "Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole)," *Bulletin de la Société Anatomique*, vol. 6, pp. 330–335, 1861.
- [2] C. Wernicke, *Der Aphasische Symptomencomplex. Eine Psychologische Studie auf Anatomischer Basis*. Breslau, Germany: Crohn und Weigert, 1874.
- [3] N. Geschwind, "Disconnexion syndromes in animals and man," *Brain*, vol. 88, no. 3, pp. 585–644, 1965.
- [4] F. A. Hayek, *The Sensory Order*. Chicago, IL, USA: Univ. Chicago Press, 1952.
- [5] T. Kohonen, *Self-Organization and Associative Memory*, 3rd ed. Berlin, Germany: Springer-Verlag, 1988.
- [6] G. M. Edelman and V. B. Mountcastle, *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain*

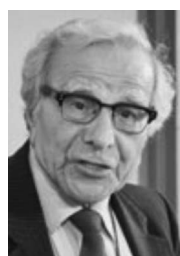


- Function. Cambridge, MA, USA: MIT Press, 1978.
- [7] D. Rumelhart and J. McClelland, *Parallel Distributed Processing: Psychological and Biological Models*. Cambridge, MA, USA: MIT Press, 1986, vol. 2.
  - [8] J. M. Fuster, "Cortex and memory: Emergence of a new paradigm," *J. Cogn. Neurosci.*, vol. 21, no. 11, pp. 2047–2072, 2009.
  - [9] J. M. Fuster and S. L. Bressler, "Cognit activation: A mechanism enabling temporal integration in working memory," *Trends Cogn. Sci.*, vol. 16, no. 4, pp. 207–218, 2012.
  - [10] J. Cowan, "Statistical mechanics of nervous nets," in *Neural Networks*. Berlin, Germany: Springer-Verlag, 1968, pp. 181–188.
  - [11] K. Friston, "Learning and inference in the brain," *Neural Netw.*, vol. 16, no. 9, pp. 1325–1352, 2003.
  - [12] S. Haykin, "Cognitive dynamic systems," *Proc. IEEE*, vol. 94, no. 11, pp. 1910–1911, Nov. 2006.
  - [13] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
  - [14] S. Haykin, "Cognitive radar: A way of the future," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 30–40, Jan. 2006.
  - [15] S. Haykin, *Cognitive Dynamic Systems: Perception-Action Cycle, Radar and Radio*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
  - [16] S. Haykin, "Cognitive dynamic systems: Radar, control, and radio," *Proc. IEEE*, vol. 100, no. 7, pp. 2095–2103, Jul. 2012.
  - [17] J. M. Fuster, *Cortex and Mind: Unifying Cognition*. New York, NY, USA: Oxford Univ. Press, 2003.
  - [18] M. I. Posner, *Cognitive Neuroscience of Attention*, 2nd ed. New York, NY, USA: Guilford, 2011.
  - [19] R. M. Neal. [Online]. Available: <http://www.cs.utoronto.ca/~radford/res-bayes-ex.html>
  - [20] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Berlin, Germany: Springer-Verlag, 2001.
  - [21] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
  - [22] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.
  - [23] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2009.
  - [24] A. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math. Dokl.*, vol. 5, pp. 1035–1038, 1963.
  - [25] A. Amiri and S. Haykin, "Improved sparse coding under the influence of perceptual attention," *Neural Comput.*, vol. 26, no. 2, pp. 377–420, 2014.
  - [26] H. Feldman and K. J. Friston, "Attention, uncertainty, and free-energy," *Front. Human Neurosci.*, vol. 4, 2010, DOI: 10.3389/fnhum.2010.00215.
  - [27] S. Haykin, *Adaptive Filter Theory*, 5th ed. Upper Saddle River, NJ, USA: Pearson, 2013.
  - [28] R. E. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.
  - [29] M. Fatemi and S. Haykin, "On reinforcement learning and Bellman's optimality in cognitive control," *IEEE Trans. Syst. Man Cybern.*, 2014.
  - [30] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
  - [31] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ, USA: Princeton Univ. Press, 1961, vol. 4.
  - [32] E. K. Miller and J. D. Cohen, "An integrative theory of prefrontal cortex function," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 167–202, 2001.
  - [33] J. M. Fuster, *The Prefrontal Cortex*, 4th ed. New York, NY, USA: Academic, 2008.
  - [34] E. Koechlin and A. Hyafil, "Anterior prefrontal function and the limits of human decision-making," *Science*, vol. 318, no. 5850, pp. 594–598, 2007.
  - [35] D. Badre, J. Hoffman, J. W. Cooney, and M. D'Esposito, "Hierarchical cognitive control deficits following damage to the human frontal lobe," *Nature Neurosci.*, vol. 12, no. 4, pp. 515–522, 2009.
  - [36] C. Azuar, P. Reyes, A. Slachevsky, E. Volle, S. Kinkingnehun, F. Kouneiher, E. Bravo, B. Dubois, E. Koechlin, and R. Levy, "Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions," *NeuroImage*, vol. 84, pp. 1053–1060, 2014.
  - [37] A. Baddeley, "Working memory," *Philosoph. Trans. Roy. Soc. Lond. B, Biol. Sci.*, vol. 302, no. 1110, pp. 311–324, 1983.
  - [38] A. J. Shackman, T. V. Salomons, H. A. Slagter, A. S. Fox, J. J. Winter, and R. J. Davidson, "The integration of negative affect, pain and cognitive control in the cingulate cortex," *Nature Rev. Neurosci.*, vol. 12, no. 3, pp. 154–167, 2011.
  - [39] J. C. Dreher, "Neural coding of computational factors affecting decision making," *Progr. Brain Res.*, vol. 202, pp. 289–320, 2012.
  - [40] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA, USA: Athena Scientific, 1995, vol. 1.
  - [41] S. Haykin, A. Zia, Y. Xue, and I. Arasaratnam, "Control theoretic approach to tracking radar: First step towards cognition," *Digital Signal Process.*, vol. 21, no. 5, pp. 576–585, 2011.
  - [42] S. Haykin, M. Fatemi, P. Setoodeh, and Y. Xue, "Cognitive control," *Proc. IEEE*, vol. 100, no. 12, pp. 3156–3169, Dec. 2012.
  - [43] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
  - [44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2012.
  - [45] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd ed. New York, NY, USA: Wiley, 2011.
  - [46] S. Haykin, Y. Xue, and P. Setoodeh, "Cognitive radar: Step toward bridging the gap between neuroscience and engineering," *Proc. IEEE*, vol. 100, no. 11, pp. 3102–3130, Nov. 2012.
  - [47] G. Mongillo, H. Shteingart, and Y. Loewenstein, "The misbehavior of reinforcement learning," *Proc. IEEE*, vol. 102, *Special Issue on Cognitive Dynamic Systems*, no. 4, Apr. 2014, DOI: 10.1109/JPROC.2014.2307022.
  - [48] T. L. Fine, *Probability and Probabilistic Reasoning for Electrical Engineering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2006.
  - [49] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
  - [50] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. New York, NY, USA: Academic, 2009.
  - [51] J. M. Fuster, *The Neuroscience of Freedom and Creativity: Our Predictive Brain*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
  - [52] A. Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science," *Behav. Brain Sci.*, vol. 36, no. 3, pp. 181–204, 2013.

## ABOUT THE AUTHORS

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

He is a Distinguished University Professor in the Faculty of Engineering, McMaster University, Hamilton, ON, Canada. For much of the past 15 years, he has focused his entire research program on learning from the human brain and applying it to a new generation of cognitive dynamic systems,



exemplified by cognitive radar, cognitive control, and cognitive radio. He is the author/coauthor of over 50 books on communication systems (analog and digital), adaptive filter theory, neural networks and learning machines, and cognitive dynamic systems.

Prof. Haykin is a Fellow of the Royal Society of Canada. He is the recipient of the Honorary Doctor of Technical Sciences (ETH, Zurich, Switzerland), the Booker Gold Medal from URSI for his outstanding contributions to radar and wireless communications, as well as many other medals and awards.

**Joaquín M. Fuster** was born in Barcelona, Spain. He studied medicine at the University of Barcelona, Barcelona, Spain. He received the Ph.D. degree in neuroscience from the University of Granada, Granada, Spain (under Prof. F. Reinoso-Suárez).

In Barcelona, Spain, and Innsbruck, Austria, he specialized in psychiatry. In 1957, he emigrated to the United States to initiate a career in neuroscience at the University of California Los Angeles (UCLA), Los Angeles, CA, USA. In 1962–1964, he worked as a Visiting Scientist at the Max-Planck Institute for Psychiatry, Munich, Germany. He is a Professor of Psychiatry and a member of the Brain Research Institute and the Semel Institute for Neuroscience and Human Behavior at the UCLA's School of Medicine. He is the author of more than 300 articles and three books. His book *The Prefrontal Cortex* is now a classic, the undisputed universal reference on its subject. The fifth edition, upgraded and expanded, is presently in preparation. Together with his colleagues, he has made several significant contributions to cognitive neuroscience. Among the most important are: 1) he showed that brain-stem reticular activation facilitates visual (tachistoscopic) attention (*Science*, 1958); 2) he pioneered intracellular records from the brain, showing a psychophysical powerfunction between visual stimulus intensity and synaptic potential amplitude in visual system (*Zeitschrift für vergleichende Physiologie*, 1965); 3) he showed reversible working-memory deficits by selective temporary cooling of cortical areas (*Brain Research*, 1970; *Experimental Neurology*, 1981); 4) with one of the earliest microelectrode drives—developed by him for the behaving animal



(*Science*, 1961)—he discovered the first “memory cells,” in the prefrontal cortex of the monkey in working memory (*Science*, 1971; *Journal of Neurophysiology*, 1973); 5) he discovered the first memory cells in inferotemporal (IT) cortex (*Science*, 1981; *Journal of Neuroscience*, 1982); 6) he showed functional interactions between IT and prefrontal cells in visual working memory (*Brain Research*, 1985); 7) he found tactile memory cells in parietal cortex (*Proceedings of the National Academy of Sciences*, 1996, 2000); 8) by using working-memory tasks of different modalities, he showed the distributed and associative nature of cortical networks in working memory (*Nature*, 2000; *Cerebral Cortex*, 2007); and 9) by neuroimaging, he has studied the structure and dynamics of memory networks in the primate (*Neuroimage*, 2005; *European Journal of Neuroscience*, 2005).

Dr. Fuster's major honors and awards include: the title of elected Member of Honor of the Spanish Royal Academy of Medicine (1997); the Signoret Prize (University of Paris) (2000); the Fyssen International Science Prize (2000); Doctor Honoris Causa, Universidad Miguel Hernández, Alicante, Spain (2003); the Goldman-Rakic Prize for Cognitive Neuroscience (NARSAD) (2006); the George Miller Prize of the Cognitive Neuroscience Society (2006); Doctor Honoris Causa, Universidad Autónoma de Madrid (2008). Recently, the University of California has instituted an endowed chair in cognitive neuroscience named after Prof. Fuster. Susan Bookheimer, a well-renown expert in neuroimaging, has been appointed the first occupant of the Fuster Chair. Related to the institution of the Chair, a suite of laboratories for cognitive neuroscience, designed by Fuster, has been opened in a new neuroscience research building in the UCLA campus.