INVITED PAPER

Neurobiologically Inspired Self-Monitoring Systems

This article outlines the neurobiological principles for living organisms that inspire self-awareness in engineered systems, using adaptive, self-monitoring robots as an exemplar for an engineered self-aware system.

By Andrea A. Chiba and Jeffrey L. Krichmar, *Senior Member IEEE*

ABSTRACT | In this article, we explore neurobiological principles that could be deployed in systems requiring self-preservation, adaptive control, and contextual awareness. We start with low-level control for sensor processing and motor reflexes. We then discuss how critical it is at an intermediate level to maintain homeostasis and predict system set points. We end with a discussion at a high level, or cognitive level, where planning and prediction can further monitor the system and optimize performance. We emphasize the information flow between these levels both from a systems neuroscience and an engineering point of view. Throughout the article, we describe the brain systems that carry out these functions and provide examples from artificial intelligence, machine learning, and robotics which include these features. Our goal is to show how biological organisms performing self-monitoring can inspire the design of autonomous and embedded systems.

KEYWORDS | Artificial intelligence; attention; adaptive control; homeostasis; machine learning; neuroscience; robotics.

I. INTRODUCTION

Neurobiology has a long history of inspiring engineering systems. The field of neural networks was derived from the architecture of the nervous system with nodes and connections that mimic neurons and synapses, respectively. Many machine learning algorithms are based on discoveries from the neuroscience of learning and memory. Robot navigation systems have been modeled after regions of the rodent brain which are important for spatial memory.

In this article, we explore neurobiological principles that monitor and regulate an organism's health and performance. Fig. 1 provides a roadmap, which we follow throughout the article. The left side of the figure lists components of the nervous system involved in primitive reflexive behavior and sensory process (bottom left, Fig. 1), the maintenance of system stability (middle left, Fig. 1), and higher level planning and control (top left, Fig. 1). The right side of Fig. 1 lists possible parallels in engineered autonomous systems.

At the lower level, the periphery and spinal cord do much of the heavy lifting with reflexive movements and rapid adjustments. Similarly, a robot would have motor controllers and drivers to handle and monitor the movement of actuators. On the sensory side, the peripheral nervous system and specialized sensors (eyes, ears, and touch receptors) handle incoming signals. These sensors do not just pass information to other brain regions. Rather, they are smart sensors that preprocess information and adapt to conditions.

At the intermediate level, subcortical systems maintain the organism's health. These portions of the nervous system regulate basic bodily functions such as hunger, thirst, heart rate, temperature control, mating, maternal/paternal instincts, defensive, and escape behaviors. These systems monitor internal organs and external conditions and then drive systems to set points appropriate for the current conditions or the organism's needs.

At the highest level, the central nervous system carries out functions that could be called "cognitive." These include attention, executive control, decision-making, navigation strategies, and planning. These functions often require learning and long-term memory. They may take time to develop and be applied. Therefore, it is critically important for the intermediate and lower levels to rapidly handle events and system

Manuscript received May 31, 2019; revised October 30, 2019 and February 5, 2020; accepted February 26, 2020. Date of publication April 6, 2020; date of current version June 18, 2020. The work of Andrea A. Chiba was supported in part by Irina Merzlyak Russell and the Chiba Laboratory (NIMH) under Grant R01MH110514-02, in part by the Wiles Laboratory, and in part by the Temporal Dynamics of Learning Center (NSF SMA) under Grant 1041755. The work of Jeffrey L. Krichmar was supported in part by the Defense Advanced Research Projects Agency (DARPA) through the Air Force Research Laboratory (AFRL) (Lifelong Learning Machines: L2M) under Contract FA8750-18-C-0103 and in part by the Air Force Office of Scientific Research (AFOSR) under Contract FA9550-19-1-0306. *(Corresponding author: Jeffrey L. Krichmar.)*

Andrea A. Chiba is with the Department of Cognitive Science, University of California at San Diego, La Jolla, CA 92093 USA.

Jeffrey L. Krichmar is with the Department of Cognitive Sciences and the Department of Computer Science, University of California at Irvine, Irvine, CA 92697-5100 USA (e-mail: jkrichma@uci.edu).

Digital Object Identifier 10.1109/JPROC.2020.2979233

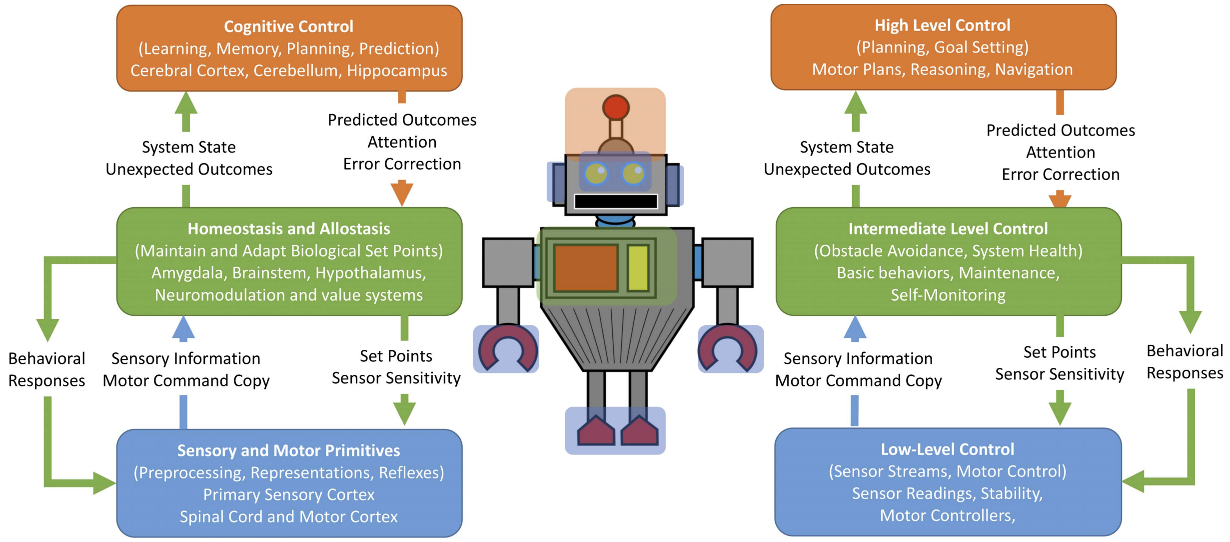

Fig. 1. *Schematic for neurobiologically inspired autonomous systems. On the left are terms and regions derived from neuroscience. On the right are terms adapted from autonomous robots, but could be applied to many embedded systems. Blue: low-level sensory processing and motor control. Green: homeostasis, maintenance, and monitoring. Orange: high-level planning, adapting, and goal-driven behavior.*

health, while the higher, "cognitive" levels plan for the future.

In the remainder of this article, we use Fig. 1 as a roadmap to discuss in more detail these levels and how the system monitors the self. In addition to covering the neurobiology behind these ideas, we will provide examples, mainly from robotics work with which we are personally familiar. By no means is this meant to be a comprehensive review. Rather, the examples are meant to illustrate our points. We hope that these ideas can inspire future autonomous systems.

II. SENSORY AND MOTOR PRIMITIVES

At the lowest level of control in Fig. 1, the organism or autonomous system needs some primitive functionality to get it out of the box. This includes actuators with motor drivers so that when a motor command for a behavioral response comes to the motor system, actuators move limbs or wheels in any desired direction and distance. A copy of that motor command is sent to the level above. In neuroscience, this is known as a motor efference copy, and it is critical for the intermediate controller to monitor the movement and make corrections if necessary [1]. Similarly, sensory systems need low-level processing so that when a stimulus, whether it is light, sound, or vibration, reaches the organism, the signal is converted into something the controller can interpret. Having some processing handled by these smart sensors and actuators reduces the load on the rest of the nervous system, which in turn saves time and energy.

A. Reflexive Behavior

Similar to biological organisms, autonomous systems need innate behaviors or reflexes to be minimally competent. For example, an organism will reflexively move away from a noxious stimulus. It may also have innate food preferences. When designing an autonomous system, engineers typically build primitive behaviors, reflexive movements, and even preferences to give it basic functionalities.

The neuroanatomist Valentino Braitenberg described a series of thought experiments for his vehicles to demonstrate a range of reflexive behaviors [2]. These vehicles had innate preference or aversion for sensory sources such as lights or sounds. The purpose of these vehicles was to provide simple lessons in neuroscience principles. For example, in the peripheral nervous system, sensory signals from one side of the body crossover to the motoneurons on the other side of the body. These contralateral connections lead to rapid, reflexive avoidance behavior. He further showed how switching the wires from sensors to ipsilateral actuators would change avoidance behavior to orienting behavior. Such an organization is found throughout the nervous system. For example, the left side of the visual cortex receives information from the right eye and part of the left eye. The left side of the motor cortex mainly drives limbs on the right side of the body. Furthermore, the type of connection makes a big difference. Changing a connection from excitatory to inhibitory will change the vehicle's behavior from avoiding to orienting and vice versa. Reflexive circuits from peripheral sensory receptors to motor neurons in the spinal cord and then to muscles are made up of these excitatory and inhibitory circuits. This organization is maintained throughout the periphery and into the central nervous system.

The spinal cord and subcortical controllers execute a number of preprogrammed behaviors [3]. These do not require learning from scratch or remembering. They are similar in vein to the idea of subsumption architecture [4], [5], which demonstrated intelligence without representation or reasoning. The subsumption architecture by Rodney Brooks′ group introduced a multitask scheduler, where different low-level sensory systems could trigger different reflexive behaviors. Arbitration between signals and prioritizing signals led to interesting behavioral repertoires. Similarly, central pattern generators (CPGs) in the spinal cord arbitrate between motor primitives [6].

B. Innate Values and Preferences

Organisms know the difference between good and bad without needing to experience and learn these preferences. Gustatory circuits have innate preferences for certain foods. Noxious stimuli are painful and lead to avoidance behavior. In general, biological organisms and artificial autonomous systems need preferences and reflexive responses, out-of-the-box, to survive. Value systems signal important events causing the organism to be aware or attend to the stimuli and trigger adaptive mechanisms that lead to remembering what to do in the future in case such an event occurs again.

In the Darwin series of brain-based devices [7], all the robots had innate values for what is good and bad. For example, different metal objects had preferred "tastes" depending on the metal's conductivity, which led to the robot learning to pick up good-tasting objects and putting down bad-tasting objects based on their associated auditory and visual cues [8]. In another set of experiments, the different reflectivity of the surface the robot traversed could also have positive or negative value, leading to orienting or fleeing behavior, respectively, [9], [10]. These innate values allowed robots to explore their environment without catastrophic failure. The exploration allowed neural networks to experience and then learn which sensory cues predicted these values and to plan accordingly. Without innate values, such learning would not be possible.

Similarly, field robots, edge devices, and the Internet of Things (IoT) should have built-in preferences to maintain connectivity and ensure system health. Building multiple innate values into the system and tying these to appropriate actions can result in adaptive behavior with minimal control policies.

III. HOMEOSTASIS AND ALLOSTASIS

At the intermediate level of control in Fig. 1, the organism or autonomous system needs to monitor external environmental conditions and internal body states to maintain system health. In biological systems, this is carried out by processes known as homeostasis and allostasis. These critical systems allow the organism to be aware of internal responses to changes in the environment (e.g., hunger, pain, and thermoregulation).

Homeostasis refers to stability through constancy and allostasis refers to predictive control of physiological conditions. Allostasis and homeostasis are complementary, that is, when predictions fail, there needs to be error correction and new set points [12]. Homeostatic mechanisms can maintain stability at these set points. These processes are governed by the hypothalamus and other subcortical regions. We will discuss later how cortical predictive control can further regulate allostasis. In Fig. 1, we suggest that allostasis and homeostasis are important for system monitoring and regulating basic behaviors. However, they are constantly sending system status to high-level controllers and are receiving context and system signals from higher levels. If predicted outcomes do not match expectations (e.g., a motor action did not result in being in an expected position), signals are sent from the intermediate controller to the higher level controller for error correction and adaptation.

A. System Health and Self-Monitoring

In the brain, the autonomic nervous system and associated physiological processes maintain system health and respond to changes [13]. There are a variety of homeostatic systems to maintain set points in the body, including thermoregulation, hunger, thirst, and protection against predators and disease. Many of these mechanisms are subconscious and reflexive, others are under voluntary control. The term "allostasis" is a process by which the body responds to stressors or changes to regain stability in the face of change. For example, a set point can change and the system must adapt and take action to restore order. The order may come in the form of a new set point that is more suitable to the current state of the environment and the current state of the organism. Matching the dynamics of the system to the dynamics of the environment or the load that is placed on the system can serve the purpose of finding a new stable state that maximizes efficiency, given the context [14], [15].

Self-monitoring and allostasis can lead to behavioral tradeoffs. The autonomic nervous system monitors whether its needs are met. If the physiological system is intent on fulfilling basic needs, such as food and sleep, it can temporarily withstand a number of problems. However, as the duration of responding in the face of unmet needs increases, the likelihood that the system will become depleted and undergo a state change also increases. For example, humans under stress might exhibit an observable allostatic change in the form of the system's compensatory down-regulation of thyroid production [16]. This lowers the system's metabolic rate, altering its energetic and restorative needs. It comes with costs, however. Low thyroid production can lead to some short-term memory deficits. The benefit of allowing the organism to continue functioning under substandard conditions, however, can outweigh the costs.

Some researchers have suggested that monitoring the system health and internal states is a step toward "self-awareness" [17], [18]. The internal representations lead to self-monitoring and can set a context for the system. On the one hand, this causes the system to adjust its actions based

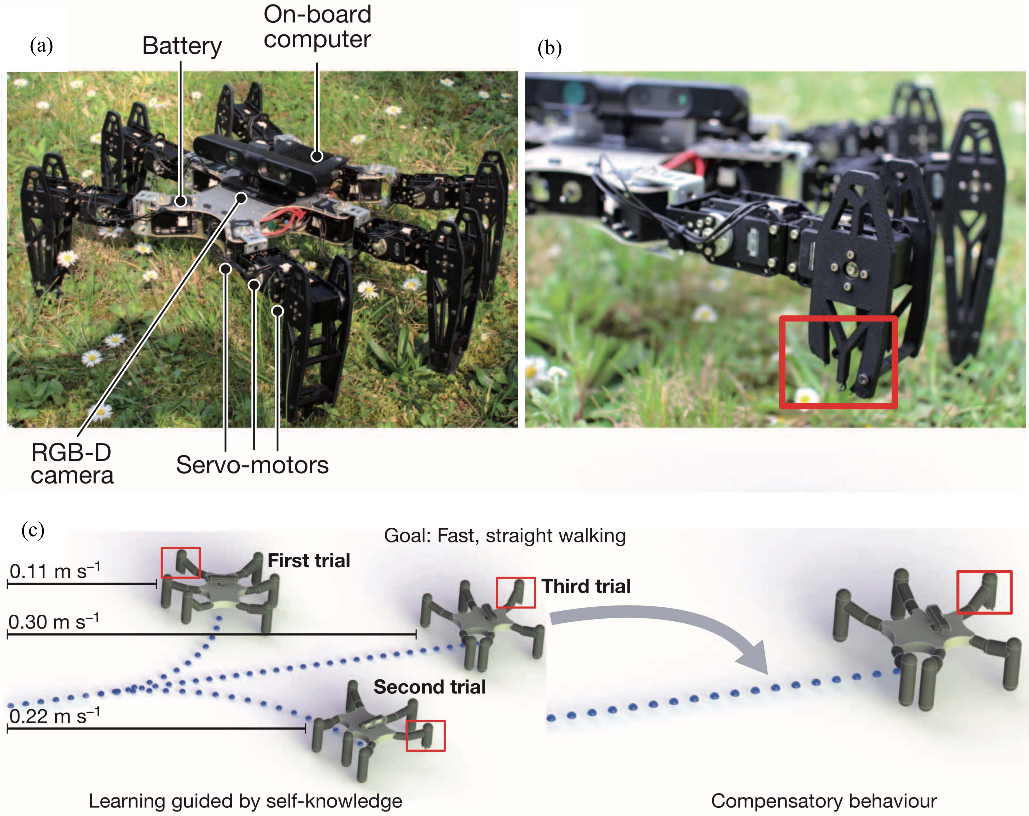

Fig. 2. ***Using an imagined trial-and-error algorithm, robots, like animals, can quickly adapt to recover from damage. (a) Undamaged, hexapod robot. RGB-D stands for red, green, blue, and depth. (b) Hexapod robot with a broken leg. (c) After damage occurs, the robot recognizes that it cannot walk fast and in a straight line. The robot tests different types of behaviors until it discovers an effective compensatory behavior. Adapted from [11]. Reprinted by permission from Springer Nature.***

on system health or needs. On the other hand, internal representations are often thought to lead to the notion of "feelings" and "awareness." In humans, monitoring of internal states often occurs below the level of awareness; however, people can bring things like feeling heart rate, respiration rate, and other signals into awareness. The concept of interoception in humans includes sensing the state of visceral organs or the internal state of the body [17]. The circuits that are proposed to support such a function (including the amygdala and insular cortex) receive input from all of the visceral organs. This ability to sense one's state is often thought to be fundamentally necessary for emotion regulation and for assessing the state of another being [18].

The concepts of allostasis and homeostasis have implications for autonomous systems. Seeking an energy source, transitioning into a power savings mode when idle, and shutting down a computer if its hardware gets too hot are examples where control modeled after homeostasis and allostasis could be advantageous. These functions do not need a central top-down control to operate. Not only do autonomous systems need to monitor their health and maintain working levels for their power consumption, sensors, and actuators, but they also need to adapt and respond to perturbations; especially if they are operating at the edge far away from power sources and support. Understanding how to maintain stability through change, as a nervous system under chronic load does, could benefit the adaptability of autonomous systems to substandard conditions or even unusually dynamic or unstable conditions, ultimately aiding survival.

B. Safety and Damage Control

Allostasis and homeostasis have similarities to Self-Integrating and Self-improving Systems or SISSY, which has been used in space systems and engineering, to guarantee safety [20]. In SISSY, the system needs to detect faults, self-protect (or safetying), and determine the minimal acceptable performance. Similarly, the hypothalamus monitors fault detection and sets the minimal acceptable performance by making new set points. Some differences may be observed, such as the nervous system and the body has built-in self-protection features and determining the appropriate performance level is dynamic and dictated by multiple signals (i.e., different brain areas). In this way, the nervous system is a more distributed variant of SISSY.

Similar to SISSY, self-monitoring and self-modeling in robotics can allow the system to recognize damage and attempt to fix or overcome an injury. For example, Cully *et al.* [11] developed a method for adapting gaits on a hexapod robot. Through self-modeling, the robot controller had a memory of potential gaits. If one or more of the robot's legs were damaged, the robot would detect the damage, "imagine" different ways of moving, and then choose the new gait it thought would work best under the new circumstances. In this way, at a low level of control, the robot monitored itself and adapted its behavior quickly without intervention (see Fig. 2).

C. Neuromodulation and Value Systems

Allostasis and homeostasis can be maintained by so-called "value systems." Organisms adapt their behavior by generating predictions that recruit value systems to maintain adequate performance and behavior. When recruited, these systems signal contextual information, trigger learning, and select actions. In the brain, these value systems are supported by neuromodulatory systems. The neuromodulatory systems are subcortical regions in the brain that have a strong influence on a number of brain areas which are considered to be involved in cognition.

Neuromodulators include dopamine, serotonin, norepinephrine, acetylcholine, and other neurochemicals that are released to a wide network of neural structures. The function of these neuromodulatory systems varies according to their actions on different target structures and receptors. For example, dopamine is thought to signal aspects of reward, saliency, novelty, invigoration, motor timing, and prediction error [4], [21], [22]. Serotonin typically contributes to feelings of well-being and security or safety. However, serotonergic producing neurons in the Raphe Nucleus undergo a paradoxical switch under conditions of high levels of threat [23], whereby these neurons then trigger threat escape behavior, including harm aversion, and might also trigger anxious states, which can lead to protective behaviors [21], [24], [25]. Under conditions of safety or lower threat, these same neurons will trigger "freezing" behavior or cessation of movement. Norepinephrine can create state changes in brain processing, signal vigilance, arousal, and under conditions of learning, track unexpected uncertainty [26], [27]. Acetylcholine is critical for inducing cortical state changes, map plasticity, sensory coding, incrementing and decrementing attention, responding to conditions of memory load, memory consolidation, attention, and tracking expected uncertainty [27]–[29]. The basal forebrain neurons, which produce acetylcholine, receive projections from all other neuromodulatory systems, perhaps serving as a final common pathway to different regions of cortex [30].

In robotics, neuromodulatory value systems can control behavior by changing the system's contextual state. For example, a robot was created to mimic rodent behavior by staying near walls or near a nest when it was anxious about an unfamiliar environment [19]. However, once it sensed that the environment was safe, the robot's curiosity increased and it explored novel objects in the middle of the environment (see Fig. 3). Simulated acetylcholine and norepinephrine allowed the robot to respond quickly to novel events and habituate to uninformative events. Increasing serotonin levels in the model led to risk-averse behavior (i.e., staying near the walls or nest), whereas increasing dopamine levels led to invigorated, curious behavior (i.e., examining objects in the middle of the environment).

For an autonomous system design, such modulation could allow a system to detect important signals from noise, and switch from one activity to another. For example, in our neural network modeling,

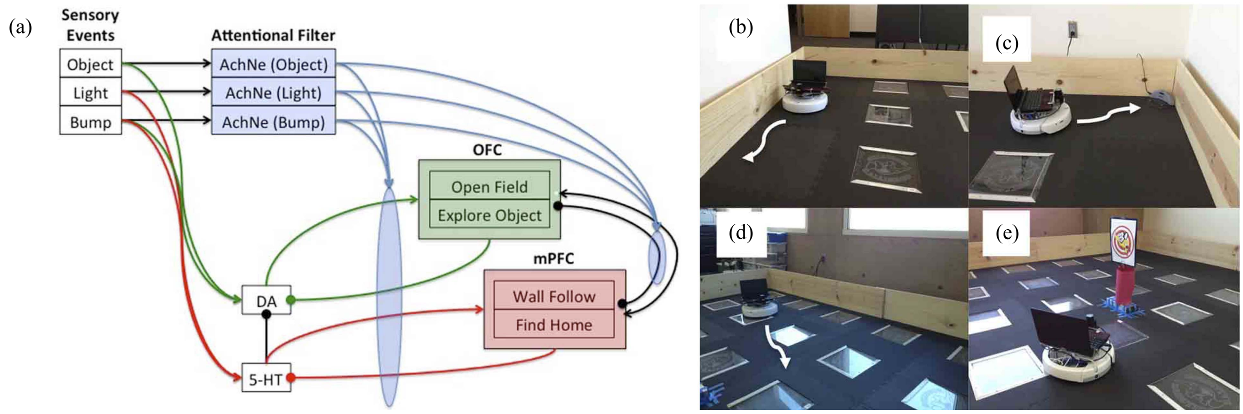

Fig. 3. *Neuromodulatory robot controller. (a) Neural network architecture to control robot behavior. Cholinergic (ACh) and noradrenergic (NE) neurons act as an attentional filter (AchNe) and the dopaminergic and serotonergic neurons (DA and 5-HT) set the level of curious or anxious behavior, respectively. The most active orbitofrontal cortex (OFC) or medial prefrontal cortex (mPFC) neuron dictated the robot's action. (b) and (c) Wall following behavior and find home (i.e., a charging station) were examples of anxious behaviors. (d) and (e) Exploring the middle of the room or approaching a novel object was an example of curious behavior. Adapted from [19] with permission.*

we have shown how neuromodulation can overcome catastrophic forgetting [31], and can lead to goal-driven perception [32]. Moreover, the noradrenergic system, which is important for one-shot learning and task switching [26], [33], [34], has important implications for solving shortcomings in deep neural networks. A strong phasic response from the noradrenergic system can clear a memory that is no longer valid, and cause rapid adaptation to new information [33]. This may be the brain's way of performing task switching and goal-directed perceptions.

IV. COGNITIVE CONTROL

At the highest level of control in Fig. 1, long-term strategies are planned and executed. Such planning requires the ability to predict outcomes and adapt when there are unexpected results. Making predictions requires the construction of internal models, which necessitates learning and memory. Since an organism or autonomous system cannot possibly monitor every signal from the environment, the higher level must prioritize which signals to receive and act upon through attention mechanisms.

The cognitive control level has similarities to cognitive architectures (for a review, see [35]). Many of these architectures include modules for attention, perception, action selection, learning and memory, reasoning, and other cognitive functions. However, most of these cognitive architectures do not consider systems-level neuroscience. Rather their goal is to extract principles from human cognitive neuroscience into modules with specific functions. What we argue for here is an approach that takes into consideration the anatomy and dynamics of the cortical and subcortical nervous systems. Moreover, we argue that all levels of our approach are closely coupled to the body. In the sections that follow, we look at many of these cognitive functions and how they monitor the self and body.

A. Predictive Control

Predictions can be generated based on prior experience and knowledge about the world, including benefits, liabilities, salience, and the statistics of the natural world. Whereas the elements of the neural system develop predictive capacities, cortical regions (frontal and parietal) that continually receive highly processed incoming sensory input and input from systems with memory or motor capacity can realize predictive coding at a pivotal level for the organism [36]. These brain areas set goals and predict outcomes [37]. Preempting reflexes, predicting value, and goal seeking can improve system performance. Moreover, the prediction is related to minimizing energy from an information theory standpoint [38]. By minimizing surprises and unanticipated events, the system can reduce energy expenditure.

Prediction requires the construction and maintenance of internal models. The brain maintains internal models for a wide range of behaviors; from motor control to language processing [39], [40]. There is evidence for neural correlates of model-based reinforcement learning in the prefrontal cortex, where an internal model is maintained to predict the value of future decisions [41]. In the rodent hippocampus, neural traces have been observed that appear to be evaluating different options before taking action [42], [43]. Prediction and inference are fundamental computations in cognitive systems [44].

Predictive models in the brain allow the organism to plan for the future and are advantageous when deliberation before action is possible. Notably, predictions and internal models in humans are prone to error in probabilistic reasoning. Whereas the propensity for human decision making "errors" are often viewed as irrational, it also remains possible that at least a subset of these errors is adaptive to survival [45]. For example, overestimating the possibility of life-threatening events might be important for survival.

These predictive strategies have been deployed in a wide range of robot applications. For example, robot controllers develop internal models to predict the movement of objects and other robots [46], [47]. In other cases, robots predict positive and negative value, which leads to maximizing exposure to positive objects and minimizing encounters with aversive objects [8], [48]. Through experience, these robots learned auditory and visual categories in an unsupervised manner. Encounters with objects caused the appropriate reflex action and value system response to associate the value and appropriate action with the object. In this way, the perception of an object would result in the value being anticipated and the action occurring earlier. Another robot modeled a predictive motor control region, known as the cerebellum, to develop "preflexes" [49]. Specifically, the robot used optic flow to predict collisions. Awkward, erratic movements due to collisions were replaced with smooth collision-free navigation through cluttered environments (see Fig. 4). Other neurobiologically inspired models of navigation build predictions of object and goal locations. Robots were able to learn pathways, and even communicate spatial trajectories and temporal references to other robots using this knowledge [50], [51].

B. Attention Systems

Given the vast number of potential stimuli available to an organism at any given moment, attention systems are necessary to reduce processing and focus responses to only those stimuli that are salient to the organism [52]–[54]. Saliency can depend on context or priorities, and attention can be reactive or predictive. Stimulus-driven or bottom-up attention causes rapid responses to salient stimuli. For example, an object (say a dog) moving in the periphery may cause a rapid shift of the eyes to focus attention on the dog and follow its movements. Goal-driven or top-down attention can cause the system to filter signals being

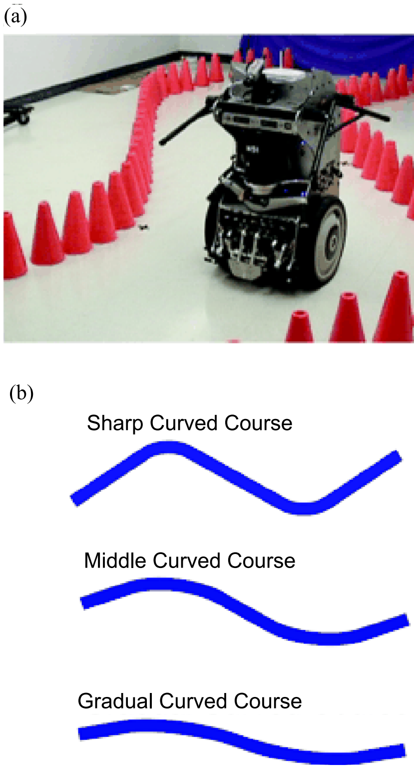

Fig. 4. ***(a) Segway Robotic Mobile Platform navigated a path dictated by orange traffic cones. An adaptive control model based on the cerebellum allowed the robot to learn smooth obstacle-free trajectories through predictive learning. (b) Layout of the different courses. Adapted from [49]. Copyright (2006) National Academy of Sciences.***

processed by the brain. For example, looking for a specific object (again, say a dog) may cause a visual search to only pay attention to objects with the size, shape, and texture of dogs. Areas such as the prefrontal cortex or the parietal cortex drive attention to features and spatial locations, respectively [52], [55].

In addition to cortical influences, neuromodulatory systems are well situated to drive attention in the cortex and other brain regions. For example, the basal forebrain cholinergic system has cortically projecting neurons that can quickly change the firing properties and the structure of firing correlations in cortex to maximize sensory coding for processing and can increase attention for learning [56], [57]. This region of the basal forebrain can result in clearer visual coding or superior auditory tracking. Another region of the basal forebrain contains neurons that project to the hippocampus to change its processing state in order to decrement the attention to irrelevant aspects of the environment. This implies that the attention system increases the signal-to-noise ratio to take in highly valued information while ignoring irrelevant information. Interestingly, the basal forebrain has temporal dynamics that might allow information to be conveyed to cortex via principles of multiplexing from information theory [58]. As such, the cortex might engage in demultiplexing in order to gain access to temporally precise information to guide attention and action.

In addition to the cholinergic system, the neuromodulator norepinephrine can rapidly switch the organism's focus of attention and induce scanning that is effective for assessing threats [59]. The locus coeruleus, which is the source of norepinephrine, has sweeping projections to the cortex, providing a mechanism by which the entire cortex can be aroused when very large state changes must be induced. Such a state change can, for example, change the perception of time, by speeding the system up enough to take in information and rapidly switch attention for advantageous decision making and survival purposes [60].

Modeling attention systems has become popular in artificial neural networks [61]. Similar to the basal forebrain, some of these artificial attention systems have an incremental component with a mask acting as a decrementing component [62]. Furthermore, ideas from how the neuromodulators acetylcholine and norepinephrine track uncertainties in the environment [27], have led to the design of goal-driven attention neural networks [32].

Being able to track the uncertainties in the world and rapidly change the focus of attention is critically important for self-monitoring and safety. Take, for example, an instance on the road in which an autonomous car might be avoiding an accident, changing states quickly would allow it to consider different sources of incoming information, switch attention to the most salient information, make adaptive decisions, and modify its actions before damage could occur.

C. Learning and Memory

A critically important aspect of humans and other animals is the ability to learn and retain information. We are able to learn over a lifetime and rapidly learn new information or skills. Learning allows us to remember facts and events of our lives and this leads to an awareness of how our past might influence the present and future. This is very different than how artificial neural networks learn and remember. Typically, artificial neural networks are trained on huge data sets for thousands to millions of training epochs. When there is new information, these networks need to be retrained and often succumb to catastrophic forgetting of old information. Moreover, slight changes to the data can cause dramatic failures [63].

The brain can offer clues on how to overcome these shortcomings in artificial lifelong learning systems. For example, the hippocampus can learn new information rapidly, and this information gets consolidated in the neocortex over time [64]. This idea of interleaved learning can overcome catastrophic forgetting. Recent results have also shown that the cortex can rapidly learn new information if it fits within a context or schema [65]–[67]. It has been shown that, in neural networks, having a schema memory

Fig. 5. *Schema network, from [31] and [68], implemented on the Toyota human support robot (HSR). The HSR retrieved objects in a breakroom and a classroom schema. Search times decreased as the HSR learned which items belong to each schema.*

can alleviate catastrophic forgetting and lead to contextual awareness, that is, taking the appropriate action depending on the situation [31], [68]. For example, schemas for rooms in a house may assist in a robot finding an object, such as a piece of fruit is typically found in the kitchen (see Fig. 5).

Hippocampal memory is also important for navigation and has neurons that encode heading, place, and path integration [69]. These spatial representations have inspired a number of robotic navigation systems [9], [70]. A recent discovery in neuroscience was a repeating pattern in animals as they move through space [71]. These neurons, known as grid cells, have inspired deep neural networks capable of navigation [72] and robot simultaneous localization and mapping (SLAM) systems that rival nonneural SLAM systems in performance [73]. In general, the hippocampus and surrounding regions have been an important inspiration for developing neurobiologically based navigation systems (for a review, see [74]). Most of these robot systems have been based on rodent experiments. Any navigation system that could come close to the rodent's capabilities would be a huge advance for robot navigation as anyone who has witnessed how well the rodent gets around complex environments can attest.

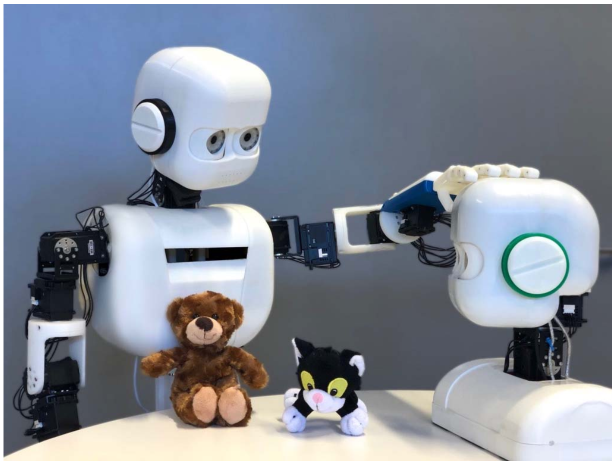

Fig. 6. *Epi is a humanoid robot developed by LUCS Robotics Group at Lund University in Sweden (https://www.lucs.lu.se/epi). It is designed to be used in developmental robotics experiments. The irises of its eyes can change color and the pupils can dilate and contract.*

D. Affective Behavior

An important part of cognition is the ability to express and recognize affect or emotions [17], [18]. For robots and intelligent agents to interact more naturally with people, they may need to have or emulate emotions [75]. The eyes can convey a wide range of emotions. Johansson and Balkenius [76] have developed a detailed model of the brain areas that control pupil dilation. This led to the development of a robot with eyes that have a strong emotional effect (see Fig. 6). Robots that seem more natural and are easier to understand through nonverbal signals may overcome the so-called "uncanny valley" and be more reliable companions.

Socially affective robots have been introduced for rat–robot interaction studies [77]. Initial studies demonstrated that rats discriminated between a social and nonsocial robot and were more likely to release a trapped robot from a cage who had helped them out of the cage in the past [78]. This suggests that rats monitor not only

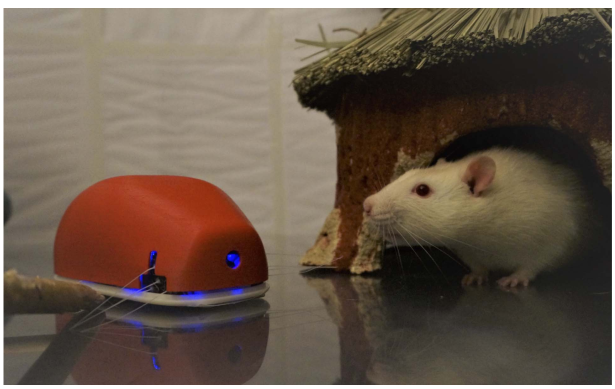

Fig. 7. *Rodent–robot interaction using the PiRat robot. Adapted from [77] with permission.*

themselves, but also their relation with other individuals. The next-generation robot, PiRat, engaged the rats and featured a control scheme capable of adapting its behavior to the state of a rat (see Fig. 7). Results showed that the rats took different trajectories according to the different behaviors of the robot. This could lead to a framework where social interaction could be studied in more controlled situations. It may also allow the robot to adapt its behavior in response to the state of another agent, which could lead to applications for robotic caretakers, assistants, or search and rescue teams.

V. CONCLUSION

In this article, we cover a range of neurobiological topics with the potential to inform self-awareness in autonomous systems, starting from innate reflexes, which allow the system to have basic competency, to cognitive functions, such as attention, memory, and social behavior. The autonomic nervous system monitors system health, keeps the organism within the operating range, and triggers system repair. Value systems, which are based on neuromodulation, can provide alerts, drive learning, and change context. Predictive coding leads to planning, goal-driven behavior, and model-based learning. Attention and memory systems have applications in computer vision systems, search systems, and navigation.

All of the aforementioned brain-inspired principles have implications for autonomous systems. The organizing principles of the nervous system, which are described here, could be applied to embedded systems, IoT, self-driving vehicles, and robots (see Fig. 1). In general, the nervous system is monitoring the self at multiple levels. It is making predictions on what to expect, and dynamically setting expectations based on environmental conditions, as well as internal needs. This functionality could lead to more autonomy and more flexibility in embedded systems. Moreover, it could realize a new class of devices and artifacts that demonstrate the intelligent and complex behavior we associate with biological systems.

Although this form of self-monitoring presented here can often be done fluidly and below the level of self-awareness, it can also be brought into awareness for heightened comprehension or decisive action. The concept that an organism can switch between self-monitoring and self-awareness also allows for the continuous processing of crucial functions. Bringing self-awareness into play can be energetically costly but it can also be beneficial. For example, self-awareness may be socially necessary for making important self-other distinctions. An animal might feel distressed by the actions of another animal. Disambiguating this feeling with respect to self or others could call attention to the fact that the other animal is in pain and requires help. This highlights the potential of this view of systems not only for building single robots but also for building functional teams.

Although what we presented here has parallels with cognitive architectures [35] and the human-like architecture for cognition and affect or H-CogAff [79], our proposed self-monitoring architecture is taken from a systems neuroscience point of view. Sloman and Chrisley [79] do discuss self-monitoring in H-CogAff, but in the context of conscious experience, which is not considered here. Rather, we take an approach grounded in experimental neuroscience. One of us is an experimental neuroscientist who studies the neural circuits that lead to behavioral repertoires. One author, Chiba, is an experimental neuroscientist who studies the neural circuits that lead to behavioral repertoires. The other author, Krichmar, is a computational neuroscientist who implements these neural circuits in devices that demonstrate the behaviors observed in animals. We believe following the neuroscience and the neurorobot examples given in this article could lead to new designs for autonomous systems. Certainly, deep learning networks [80] and deep reinforcement learning algorithms [81] would gain by incorporating more realistic neuroanatomy and neurodynamics into their models. Therefore, it is our hope that this article can inspire innovations in the design of autonomous and embedded systems. ■

Acknowledgment

Jeffrey L. Krichmar would like to thank the Cognitive Anteater Robotics Laboratory (CARL) for their inspiration and innovation. Andrea A. Chiba would like to thank Irina Merzlyak, Christopher Cordova, Pai-Kai Huang, Victor Minces, Eric Leonardis, Laleh Quinn, and other members of her laboratory for inspiring discussions and contributions.

REFERENCES

[1] R. Shadmehr and S. Wise, *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. Cambridge, MA, USA: Bradford Book, 2004.
[2] V. Braitenberg, *Vehicles, Experiments in Synthetic Psychology*. Cambridge, MA, USA: MIT Press, 1986.
[3] L. W. Swanson, "Cerebral hemisphere regulation of motivated behavior," *Brain Res.*, vol. 886, nos. 1–2, pp. 113–164, 2000.
[4] R. A. Brooks, "Intelligence without representation," *Artif. Intell.*, vol. 47, nos. 1–3, pp. 139–159, Jan. 1991.
[5] R. A. Brooks, "Intelligence without reason," in *Proc. 12th Int. Joint Conf. Artif. Intell.*, vol. 1. San Mateo, CA, USA: Morgan Kaufmann, 1991, pp. 569–595.
[6] C. L. Barkan and E. Zornik, "Feedback to the future: Motor neuron contributions to central pattern generator function," *J. Exp. Biol.*, vol. 222, no. 16, Aug. 2019, Art. no. jeb193318.
[7] G. M. Edelman, "Learning in and from brain-based devices," *Science*, vol. 318, no. 5853, pp. 1103–1105, Nov. 2007.
[8] J. L. Krichmar, "Machine psychology: Autonomous behavior, perceptual categorization and conditioning in a brain-based device," *Cerebral Cortex*, vol. 12, no. 8, pp. 818–830, Aug. 2002.
[9] J. L. Krichmar, D. A. Nitz, J. A. Gally, and G. M. Edelman, "Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 6, pp. 2111–2116, Feb. 2005.
[10] A. K. Seth, J. L. McKinstry, G. M. Edelman, and J. L. Krichmar, "Active sensing of visual and tactile stimuli by brain-based devices," *Int. J. Robot. Autom.*, vol. 19, no. 4, pp. 222–238, 2004.
[11] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, May 2015.
[12] J. Schulkin and P. Sterling, "Allostasis: A brain-centered, predictive mode of physiological regulation," *Trends Neurosci.*, vol. 42, no. 10, pp. 740–752, Oct. 2019.
[13] W. Janig, *Integrative Action of the Autonomic Nervous System: Neurobiology of Homeostasis*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[14] D. S. Ramsay and S. C. Woods, "Clarifying the roles of homeostasis and allostasis in physiological regulation.," *Psychol. Rev.*, vol. 121, no. 2,

pp. 225–247, 2014.
[15] B. S. McEwen and J. C. Wingfield, "What is in a name? Integrating homeostasis, allostasis and stress," *Hormones Behav.*, vol. 57, no. 2, pp. 105–111, Feb. 2010.
[16] A. Chatzitomaris *et al.*, "Thyroid allostasis–adaptive responses of thyrotropic feedback control to conditions of strain, stress, and developmental programming," *Frontiers Endocrinol.*, vol. 8, p. 163, Jul. 2017.
[17] A. D. (Bud) Craig, "How do you feel—now? The anterior insula and human awareness," *Nature Rev. Neurosci.*, vol. 10, no. 1, pp. 59–70, Jan. 2009.
[18] A. Damasio, "Feelings of emotion and the self," *Ann. New York Acad. Sci.*, vol. 1001, no. 1, pp. 253–261, Oct. 2003.
[19] J. L. Krichmar, "A neurorobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior," *Frontiers Neurorobotics*, vol. 7, pp. 1–17, 2013.
[20] K. Bellman, "What reasonable guarantees can we make for a SISSY system," in *Proc. IEEE 3rd Int. Workshops Found. Appl. Self Syst. (FAS W)*, Sep. 2018, pp. 231–233.
[21] Y.-L. Boureau and P. Dayan, "Opponency revisited: Competition and cooperation between dopamine and serotonin," *Neuropsychopharmacology*, vol. 36, no. 1, pp. 74–97, Jan. 2011.
[22] P. Redgrave and K. Gurney, "The short-latency dopamine signal: A role in discovering novel actions?" *Nature Rev. Neurosci.*, vol. 7, no. 12, pp. 967–975, Dec. 2006.
[23] C. Seo *et al.*, "Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode," *Science*, vol. 363, no. 6426, pp. 538–542, Feb. 2019.
[24] M. J. Crockett, L. Clark, G. Tabibnia, M. D. Lieberman, and T. W. Robbins, "Serotonin modulates behavioral reactions to unfairness," *Science*, vol. 320, no. 5884, p. 1739, Jun. 2008.
[25] M. Tops, S. Russo, M. A. S. Boksem, and D. M. Tucker, "Serotonin: Modulator of a drive to withdraw," *Brain Cogn.*, vol. 71, no. 3, pp. 427–436, Dec. 2009.
[26] S. Bouret and S. J. Sara, "Network reset: A simplified overarching theory of locus coeruleus noradrenaline function," *Trends Neurosciences*, vol. 28, no. 11, pp. 574–582, Nov. 2005.
[27] A. J. Yu and P. Dayan, "Uncertainty, neuromodulation, and attention," *Neuron*, vol. 46, no. 4, pp. 681–692, May 2005.
[28] M. E. Hasselmo and J. McGaughy, "High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation," *Prog. Brain Res.*, vol. 145, pp. 207–231, Jan. 2004.
[29] M. G. Baxter and A. A. Chiba, "Cognitive functions of the basal forebrain," *Current Opinion Neurobiol.*, vol. 9, no. 2, pp. 178–183, Apr. 1999.
[30] M. R. Gielow and L. Zaborszky, "The input-output relationship of the cholinergic basal forebrain," *Cell Rep.*, vol. 18, no. 7, pp. 1817–1830, Feb. 2017.
[31] T. Hwu and J. L. Krichmar, "A neural model of schemas and memory encoding," *Biol. Cybern.*, Nov. 2019.
[32] X. Zou, S. Kolouri, P. K. Pilly, and J. L. Krichmar, "Neuromodulated attention and goal-driven perception in uncertain domains," *Neural Netw.*, vol. 125, pp. 56–69, May 2020.
[33] S. L. Grella *et al.*, "Locus coeruleus phasic, but not tonic, activation initiates global remapping in a familiar environment," *J. Neurosci.*, vol. 39, no. 3, pp. 445–455, Jan. 2019.
[34] A. Wagatsuma, T. Okuyama, C. Sun, L. M. Smith, K. Abe, and S. Tonegawa, "Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 2, pp. E310–E316, Jan. 2018.
[35] I. Kotseruba and J. K. Tsotsos, "A review of 40 years of cognitive architecture research: Core cognitive abilities and practical applications," 2016, *arXiv:1610.08602*. [Online]. Available: http://arxiv.org/abs/1610.08602
[36] A. Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science," *Behav. Brain Sci.*, vol. 36, no. 3, pp. 181–204, Jun. 2013.
[37] P. Dayan and N. D. Daw, "Decision theory, reinforcement learning, and the brain," *Cognit., Affect., Behav. Neurosci.*, vol. 8, no. 4, pp. 429–453, Dec. 2008.
[38] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, Feb. 2010.
[39] R. Shadmehr and J. W. Krakauer, "A computational neuroanatomy for motor control," *Exp. Brain Res.*, vol. 185, no. 3, pp. 359–381, Mar. 2008.
[40] G. Hickok, J. Houde, and F. Rong, "Sensorimotor integration in speech processing: Computational basis and neural organization," *Neuron*, vol. 69, no. 3, pp. 407–422, Feb. 2011.
[41] J. Gläscher, N. Daw, P. Dayan, and J. P. O'Doherty, "States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning," *Neuron*, vol. 66, no. 4, pp. 585–595, May 2010.
[42] B. E. Pfeiffer and D. J. Foster, "Hippocampal place-cell sequences depict future paths to remembered goals," *Nature*, vol. 497, no. 7447, pp. 74–79, May 2013.
[43] B. Schmidt, A. Papale, A. D. Redish, and E. J. Markus, "Conflict between place and response navigation strategies: Effects on vicarious trial and error (VTE) behaviors," *Learn. Memory*, vol. 20, no. 3, pp. 130–138, Feb. 2013.
[44] M. Richert, D. Fisher, F. Piekniewski, E. M. Izhikevich, and T. L. Hylton, "Fundamental principles of cortical computation: Unsupervised learning with prediction, compression and feedback," 2016, *arXiv:1608.06277*. [Online]. Available: http://arxiv.org/abs/1608.06277
[45] J. Steiner and C. Stewart, "Perceiving prospects properly," *Amer. Econ. Rev.*, vol. 106, no. 7, pp. 1601–1631, Jul. 2016.
[46] J.-C. Park, J. H. Lim, H. Choi, and D.-S. Kim, "Predictive coding strategies for developmental neurorobotics," *Frontiers Psychol.*, vol. 3, p. 134, May 2012.
[47] S. Murata, Y. Yamashita, H. Arie, T. Ogata, S. Sugano, and J. Tani, "Learning to perceive the world as probabilistic or deterministic via interaction with others: A neuro-robotics experiment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 830–848, Apr. 2017.
[48] B. R. Cox and J. L. Krichmar, "Neuromodulation as a robot controller: A brain inspired design strategy for controlling autonomous robots," *IEEE Robot. Autom. Mag.*, vol. 16, no. 3, pp. 72–80, Sep. 2009.
[49] J. L. McKinstry, G. M. Edelman, and J. L. Krichmar, "A cerebellar model for predictive motor control tested in a brain-based device," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 9, pp. 3387–3392, Feb. 2006.
[50] S. Heath, D. Ball, and J. Wiles, "Lingodroids: Cross-situational learning for episodic elements," *IEEE Trans. Cogn. Develop. Syst.*, vol. 8, no. 1, pp. 3–14, Mar. 2016.
[51] S. Heath, R. Schulz, D. Ball, and J. Wiles, "Lingodroids: Learning terms for time," in *Proc. IEEE Int. Conf. Robot. Autom.*, Saint Paul, MN, USA, May 2012, pp. 1862–1867.
[52] J. H. R. Maunsell and S. Treue, "Feature-based attention in visual cortex," *Trends Neurosci.*, vol. 29, no. 6, pp. 317–322, Jun. 2006.
[53] S. K. A. L. G. Ungerleider, "Mechanisms of visual attention in the human cortex," *Annu. Rev. Neurosci.*, vol. 23, no. 1, pp. 315–341, Mar. 2000.
[54] E. I. Knudsen, "Fundamental components of attention," *Annu. Rev. Neurosci.*, vol. 30, no. 1, pp. 57–78, Jul. 2007.
[55] M. Husain and P. Nachev, "Space and the parietal cortex," *Trends Cogn. Sci.*, vol. 11, no. 1, pp. 6–30, 2007.
[56] V. Minces, L. Pinto, Y. Dan, and A. A. Chiba, "Cholinergic shaping of neural correlations," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 22, pp. 5725–5730, May 2017.
[57] L. Quinn and A. A. Chiba, "Basal forebrain and memory," in *Learning and Memory: A Comprehensive Reference*. New York, NY, USA: Academic, 2008, pp. 281–301.
[58] D. Tingley, A. S. Alexander, L. K. Quinn, A. A. Chiba, and D. Nitz, "Multiplexed oscillations and phase rate coding in the basal forebrain," *Sci. Adv.*, vol. 4, no. 8, Aug. 2018, Art. no. eaar3230.
[59] M. Usher, "The role of locus coeruleus in the regulation of cognitive performance," *Science*, vol. 283, no. 5401, pp. 549–554, Jan. 1999.
[60] T. W. Suzuki, J. Kunimatsu, and M. Tanaka, "Correlation between pupil size and subjective passage of time in non-human primates," *J. Neurosci.*, vol. 36, no. 44, pp. 11331–11337, Nov. 2016.
[61] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 30, 2017, pp. 5998–6008.
[62] J. Zhang, S. A. Bargal, Z. Lin, J. Brandt, X. Shen, and S. Sclaroff, "Top-down neural attention by excitation backprop," *Int. J. Comput. Vis.*, vol. 126, no. 10, pp. 1084–1102, Oct. 2018.
[63] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–11.
[64] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory," *Psychol. Rev.*, vol. 102, no. 3, pp. 419–457, 1995.
[65] D. Tse *et al.*, "Schemas and memory consolidation," *Science*, vol. 316, no. 5821, pp. 76–82, Apr. 2007.
[66] D. Tse *et al.*, "Schema-dependent gene activation and memory encoding in neocortex," *Science*, vol. 333, no. 6044, pp. 891–895, Aug. 2011.
[67] M. T. R. van Kesteren, S. F. Beul, A. Takashima, R. N. Henson, D. J. Ruiter, and G. Fernández, "Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent," *Neuropsychologia*, vol. 51, no. 12, pp. 2352–2359, Oct. 2013.
[68] T. Hwu, H. Kashyap, and J. Krichmar, "A neurobiological schema model for contextual awareness in robotics," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2020, pp. 1–8.
[69] D. Derdikman and E. I. Moser, "A manifold of spatial maps in the brain," *Trends Cogn. Sci.*, vol. 14, no. 12, pp. 561–569, 2010.
[70] P. Gaussier *et al.*, "Merging information in the entorhinal cortex: What can we learn from robotics experiments and modeling?" *J. Exp. Biol.*, vol. 222, Feb. 2019, Art. no. jeb186932.
[71] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," *Nature*, vol. 436, no. 7052, pp. 801–806, Aug. 2005.
[72] A. Banino *et al.*, "Vector-based navigation using grid-like representations in artificial agents," *Nature*, vol. 557, no. 7705, pp. 429–433, May 2018.
[73] M. J. Milford, J. Wiles, and G. F. Wyeth, "Solving navigational uncertainty using grid cells on robots," *PLoS Comput. Biol.*, vol. 6, no. 11, 2010, Art. no. e1000995.
[74] P. J. Zeno, S. Patel, and T. M. Sobh, "Review of neurobiologically based mobile robot navigation system research performed since 2000," *J. Robot.*, vol. 2016, Sep. 2016, Art. no. 8637251.
[75] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robot. Autonom. Syst.*, vol. 58, no. 3, pp. 322–332, Mar. 2010.
[76] B. Johansson and C. Balkenius, "A computational model of pupil dilation," *Connection Sci.*, vol. 30, no. 1, pp. 5–19, Jan. 2018.
[77] S. Heath *et al.*, "PiRat: An autonomous framework for studying social behaviour in rats and robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 7601–7608.
[78] L. K. Quinn *et al.*, "When rats rescue robots," *Animal Behav. Cogn.*, vol. 5, no. 4, pp. 368–379, Nov. 2018.
[79] A. Sloman and R. Chrisley, "Virtual machines and consciousness," *J. Consciousness Stud.*, vol. 10, nos. 4–5, pp. 133–172, 2003.
[80] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[81] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

ABOUT THE AUTHORS

Andrea A. Chiba is currently a Professor with the Department of Cognitive Science and in the Program of Neuroscience at the University of California at San Diego, La Jolla, CA, USA. Her laboratory, Chiba Laboratory, is focused on gaining an understanding of the neural systems and principles underlying aspects of learning, memory, affect, and attention, with an emphasis on neural plasticity. Her laboratory team is highly interdisciplinary, using a variety of neurobiological, neurochemical, neurophysiology, computational, robotic, and behavioral techniques. She has been active in the U.S. Brain Initiative. She is also the Co-Director and the Founding Science Director of the Temporal Dynamics of Learning Center, an NSF Science of Learning Center at the University of California, San Diego.

Dr. Chiba and her team were awarded an NSF BRAIN Initiative Award for their work on the neural basis of prosociality and interoception.

Jeffrey L. Krichmar (Senior Member, IEEE) received the B.S. degree in computer science from the University of Massachusetts at Amherst, Amherst, MA, USA, in 1983, the M.S. degree in computer science from The George Washington University, Washington, DC, USA, in 1991, and the Ph.D. degree in computational sciences and informatics from George Mason University, Fairfax, VA, USA, in 1997.

Currently, he is a Professor with the Department of Cognitive Sciences and the Department of Computer Science, University of California at Irvine, Irvine, CA, USA. He has nearly 20 years' experience in designing adaptive algorithms, creating neurobiologically plausible network simulations, and constructing brain-based robots whose behavior is guided by neurobiologically inspired models. His research interests include neurorobotics, embodied cognition, biologically plausible models of learning and memory, neuromorphic applications and tools, and the effect of neural architecture on neural function.