Symbolic Models and Emergent Models: A Review

Juyang Weng, *Fellow, IEEE*

***Abstract*—There exists a large conceptual gap between symbolic models and emergent models for the mind. Many emergent models work on low-level sensory data, while many symbolic models deal with high-level abstract (i.e., action) symbols. There has been relatively little study on intermediate representations, mainly because of a lack of knowledge about how representations fully autonomously emerge inside the closed brain skull, using information from the exposed two ends (the sensory end and the motor end). As reviewed here, this situation is changing. A fundamental challenge for emergent models is abstraction, which symbolic models enjoy through human handcrafting. The term abstract refers to properties disassociated with any particular form. Emergent abstraction seems possible, although the brain appears to never receive a computer symbol (e.g., ASCII code) or produce such a symbol. This paper reviews major agent models with an emphasis on representation. It suggests two different ways to relate symbolic representations with emergent representations: One is based on their categorical definitions. The other considers that a symbolic representation corresponds to a brain's outside behaviors observed and handcrafted by other outside human observers; but an emergent representation is inside the brain.**

***Index Terms*—Agents, attention, brain architecture, complexity, computer vision, emergent representation, graphic models, mental architecture, neural networks, reasoning, regression, robotics, speech recognition, symbolic representation, text understanding.**

I. Introduction

OVER 55 years ago, Alan M. Turing [1] raised a well-known question: "Can machines think?" However, brain-like thinking has to do with the type of the representation and the nature of the architecture that the machines employ. As the field of artificial intelligence (AI) is inspired by human intelligence, different representations and agent architectures are all inspired by the brain to different degrees. A grand challenge is to understand the brain's internal representation which tells the working of intelligence inside the brain. Meeting this grand challenge seems necessary to enable machines to reach human level intelligence.

Meeting this grand challenge seems also necessary to enable each human to truly understand himself scientifically. A solution to this grand challenge is expected to lead to answers for many important questions. For example, to what degree human brains, primate brains, mammal brains, and vertebrate brains share the same set of principles? Why do different brains show such a wide variety of behaviors? What is developmental science? What does the developmental science tell us about ways to improve the quality of human lives in different parts of the world?

Representation and mental architecture are two tightly intertwined issues for both natural intelligence and artificial intelligence. Since the early production systems in the early 70s [2], [3] there has been an increasing flow of research on cognitive architectures. Allport [4] reviewed studies on architectures of attention and control. Langley *et al.* [5] provided an overview of cognitive architectures, mainly of the symbolic type. Orebäck & Christensen [6] reviewed a few architectures of mobile robots. Recently, in a special issue on autonomous mental development, Vernon *et al.* [7] presented a review for cognitive system architectures, with an emphasis on key architectural features that systems capable of autonomous development of mental capabilities should exhibit. Barsalou [8] gave a review about grounded cognition from the psychological points of view, emphasizing brain's modal system for perception (e.g., vision, audition), action (e.g., movement), and introspection (e.g., affect). It contrasts grounded cognition with traditional views that cognition arises from computation on amodal symbols inside a modular system. It addresses a question: "does the brain contain amodal symbols?"

This review does not mean to repeat those reviews, but rather to bring up representative models for the emphasis on the types of representation and the issue of emergence. Here I raise two questions:

> Does the brain contain computer *symbols* at all in its internal representations? Why is fully autonomous *emergence* necessary for intelligence, natural and artificial?

I argue through this review that the brain does not seem to contain any computer symbol at all in its internal representations for the extra-body environment (i.e., outside the body), and its internal representations seem all emergent and are *modal* in terms of the information origin about the extra-body environment (e.g., vision, audition, motor, glands). Incrementally dealing with unexpected new extra-body environments through real-time fully autonomous emergence inside the brain is an essence of intelligence, both natural and artificial.

There is a lack of reviews of cognitive architectures that cover the subjects of perception, cognition, and decision-making in terms of computation. A large number of cognitive architectures do not address perception or at least do not emphasize it. However, the brain spends over 50% of the cortical areas for visual information processing [9], [10]. An illuminating fact is that different cortical areas use the same 6-layer laminar structure to deal with very different signal processing tasks: vision, audition, touch, association, decision making, and motor control ([11], pp. 327–329)—indicating that we should not isolate abstract decision making from grounded sensory processing.

Manuscript received September 17, 2010; revised March 29, 2011; accepted April 26, 2011. Date of publication June 09, 2011; date of current version March 13, 2012.

The author is with the Department of Computer Science and Engineering, Cognitive Science Program, and Neuroscience Program, Michigan State University, East Lansing, MI 48824 USA (e-mail: see http://www.cse.msu.edu/~weng/).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAMD.2011.2159113

It seems that the way perception, cognition, and motor control are conducted in the brain is different from what many existing agent architectures have modeled. Grounded autonomous mental development is necessary for understanding major limitations of existing agent architectures that do not address the development of sensory systems (e.g., vision) as an integral part.

This paper will first review major agent architectures including the subjects of perception, cognition, and movement. This scope is wider than the traditional subject of symbolic cognitive architectures by including perception. Representation is the central theme of this review. Further, this review will emphasize the relationships among different types of agent architectures, instead of providing a long list of existing architectures with the features from each, so that the reader can see fundamental conceptual differences.

In the remainder of this paper, Section II outlines the framework of autonomous mental development (AMD) and discusses its necessity for understanding natural intelligence and for strong machine intelligence. Section III reviews symbolic models. Section IV overviews emergent models. Section V presents the gap between the brain-mind and existing frameworks. Section VI provides some concluding remarks.

II. AMD: A Cross-Disciplinary Subject

The human central nervous system (CNS) demonstrates a wide variety of capabilities. We collectively call them mental capabilities. They can be divided into four categories, perception, cognition, movement, and motivation. The new field of autonomous mental development aims to: 1) understand human mental development; and 2) enable machines to autonomously develop their mental skills.

AMD includes two scopes—the development of a *single* mind and the department of *multiple* minds as a society. The principles of developing a single mind might be of great value for studying the science for social development, human and robotic.

A. AMD Directly Related Research Fields

Research on how the mind works can be categorized in different ways. The following is a way proposed by Weng & McClelland [12].

1) **Explain mind behaviors**. Study human and animal behaviors under exposure to stimuli. Much of the research in traditional psychology falls into this category. This category alone is insufficient. It seems to have also led some researchers to attribute mind behaviors to something other than clearly understandable computation. For example, Penrose [13], [14], facing fundamental problems of mathematic logic, attributed mind behaviors (e.g., consciousness) to quantum computation.
2) **Model mind behaviors**. Computationally model behaviors under stimuli expositions and verify the resulting models using studies in category 1). This level of modeling does not necessarily take into account how the brain works and how the brain-mind develops from experience. Much of the productive research in AI and many recent computational studies in psychology belong to this category.
   This category alone appears also insufficient, as it has misled researchers to handcraft static symbolic models which we will discuss below. The resulting machines cannot override the static symbols and they become brittle in the real world [e.g., the symbolic hyper-graph of artificial general intelligence (AGI) [15]].
3) **Model the brain**. Computationally model how the brain works, at different scales: the cortex, the circuits, and the neurons. This level of modeling does not necessarily take into account how the brain develops from experience. Much of the research in neuroscience belong to this category, although many existing studies focus on lower levels.
   This category alone appears also insufficient. It has misled researchers to handcraft static concept boundaries in the brain, resulting in *symbolic models* (e.g., statically consider the primary visual area, V1, as edge detectors, e.g., George & Hawkins 2009 [16], or using the static Gabor filters to model dynamic V1 processing [17], [18]).
4) **Modeling brain development**. Computationally model how the brain develops, at different scales: the brains, the cortex, the circuits, and the neurons. This deeper level of modeling takes into account not only how the brain works now but also how the brain developed its ways of work in the past. Much of the research in AMD belongs to this category, although this field includes also research that does not emphasize support from brain facts (e.g., incremental learning).
   A natural question is that whether humans can design the functions of the "genome" program well enough to approach human-level intelligence without actually simulating the extremely expensive evolution of the "genome" itself. Since the cost for evolving the brain-part of the human genome is extremely high, this category seems our best hope for human level performance.

Researchers are increasingly aware of the necessity of modeling brain development. It is not only for understanding how the brain-mind works, but also for solving many bottleneck problems in AI. In the following, we further discuss such computational problems.

B. Developmental Program: Task Nonspecificity

By definition, an agent is something that senses and acts. A robot is an agent, so is a human. In the early days of AI, smart systems that caught the general public's imagination were programmed by a set of task-specific rules. The field of AI moved beyond that early stage when it started the more systematic agent methodology [19], although an agent is still a task-specific machine.

1) Task Nonspecificity: Not until the NSF and DARPA funded Workshop on Development and Learning (WDL) in April 2000 [20], [21] had the concept of the *task-nonspecific* developmental program caught the attention of many researchers. Let us first explicitly define the concept of task-nonspecificity for the purpose of this review:

Definition 1 (Task Nonspecificity): An agent is task nonspecific if: 1) during the programming phase, its programmer is not given the set of tasks that the agent will end up learning; 2)

during the learning phase, the agent incrementally learns various task execution skills from interactions with the environment using its sensors and effectors; and 3) during the task execution phase, at any time the agent autonomously figures out what tasks should be executed from the cues available from the environment.

For example, a teacher says "to address D!," the robot figures out that the next goal is to go to the address D. On its way to the address D, the robot must figure out what to do, such as look for traffic signs and avoid obstacles. After reaching the address D, it figures out that it is time to stop and report. All the goals and subgoals are acquired from the environment—outside and inside the brain.

Therefore, the *task nonspecificity* does not mean that different instances of a framework (e.g., a type of neural network) can be used for different tasks through a manual adaptation of the framework. Such an agent is still a single-task agent. Furthermore, *skill sharing* and *scaffolding* are both implied by the incremental learning mode required by the above definition. Skill sharing means each skill is useful for many tasks. Scaffolding means that early-learned skills assist the learning of later more complex skills.

By the end of WDL, the concept of *task-nonspecific* developmental program had not been well accepted by leading developmental psychologists like Esther Thelen. She sent to me *The Ontogeny of Information* by Susan Oyama [22] to support that the genomes only provide "constraints". Not being computer scientists and having seeing only some primitive artificial networks, Susan Oyama and Esther Thelen thought that a computer program is always too "rigid" to regulate the developmental process for an entity as complex as brain. Oyama wrote ([22] p. 72):

> *If the processes are "programmed," then all biological events are programmed, and the concept of program cannot be used to distinguish the innate from the acquired, the adaptive from the nonadaptive, the inevitable from the mutable.*

However, biologists had no problem with calling "Epigenetic programming" [23]–[25] even then, while they described body development. Of course, since we insist that tasks are not given, a developmental program for the brain-mind only regulates (i.e., constrains) the developmental process of the brain-mind while the brain interacts with the real physical world. Development is certainly not a totally "programed" process, since it is greatly shaped by the environment. This is true for the brain and the body.

2) Autonomous Mental Development: The human genome is a *developmental program* [25], [20], [26]. It seems that the developmental program embedded in the human genome [24] is task-nonspecific. Such a developmental program enables the growth of the human brain (and body) according to a biological architecture [27], [28], such as the six-layer laminar architecture of the cerebral cortex and the wiring under normal experience. Such a development program is species specific, sensor specific, and effector specific in the sense of providing processing areas with default sensor-area pathways, but it is not feature specific (handcrafted features). Further, the brain developed under internally generated spontaneous activities before birth enables the development of inborn reflexes [29], which are important for the early survival of the young individual right after the birth. However, the developmental program of a human is task nonspecific, as the human adult can perform many tasks, including tasks that his parents have never performed.

Early developmental programs were inspired by general-purpose learning using neural networks. Cresceptron by Weng *et al.* [30], [31], [32] internally autonomously grows (i.e., develops) a network. It seemed the first developmental network for learning to recognize general objects directly from natural complex backgrounds through interactions with human operators. It appeared also the first that segments recognized objects from natural complex backgrounds. By the mid 1990s, connectionist cognitive scientists had started the exploration of the challenging domain of development (McClelland [33], Elman *et al.* [34], Quartz & Sejnowski [35]) emphasizing ideas such as growing a network from small to large [36], and the nonstationarity of the development process [35].

The term "connectionist" has been misleading, diverting attention to only network styles of computation that do not address how the internal representations emerge without human programmer's knowledge about tasks. Furthermore, the term "connectionist" has *not* been very effective to distinguish (emergent) brain-like networks from (symbolic) networks such as Bayesian networks (also called belief nets by Peal [37] and graphic models by many) which use a web of probabilities but each network "skeleton" (base framework) is handcrafted, symbolic, and static. Jordan & Bishop [38] used "neural networks" to explicitly name symbolic graphical models. The long short-term memory (LSTM) by Hochreiter & Schmidhuber 1997 [39], CLARION by Sun *et al.* [40], [41], the hierarchical temporal memory (HTM) by George & Hawkins [16], the symbolic network scheme proposed by Albus [42], the symbolic Bayesian networks reviewed by Tenenbaum *et al.* [43] are some examples among many. Lee & Mumford [44] used a vague, but intuitive term "feature" to refer to each symbolic variable during their use of the Bayesian rule for modeling cortex computation. Such a misleading "connectionist" emphasis has allowed symbolic representations to be misinterpreted as brain-liky representations. I argue that such symbolic representations are fundamentally far from brain-like representations and, therefore, their computation is far from brain-like computation.

I propose instead to use a more clear concept *emergent model* which uses representations that fully autonomously emerge—not allowing human handcrafting or twisting after the human knows the tasks to be performed. A connectionist model does not imply the use of a fully emergent representation (see the definitions below). Connectionist computations can occur in handcrafted symbolic networks (e.g., symbolic Bayesian nets). I argue that the brain is emergent but a symbolic model cannot be.

Some researchers may think that a "mixed approach"—mixing autonomous emergence in some part with a static handcrafted design for other part—is probably most practical for machines (e.g., mixed approaches in a survey by Asada *et al.* [45]). They thought that fully autonomous development after birth—from a newborn brain to an adult brain—seems unlikely necessary for computers and robots at least at the current stage

TABLE I
A COMPARISON OF APPROACHES TO ARTIFICIAL INTELLIGENCE AND TO MODELING THE MIND

Approach	Species architecture	Extra-body concepts	Agent behaviors	Representation	Task specific
Knowledge-based	Model	Model	Model	Symbolic	Yes
Learning-based	Model	Parametrically model	Model	Symbolic	Yes
Behavior-based	Model	Avoid modeling	Model	Symbolic	Yes
Genetic	Genetic search	Parametrically model	Model	Symbolic	Yes
Developmental	Parametrically model	Avoid modeling	Minimize modeling	Emergent	No

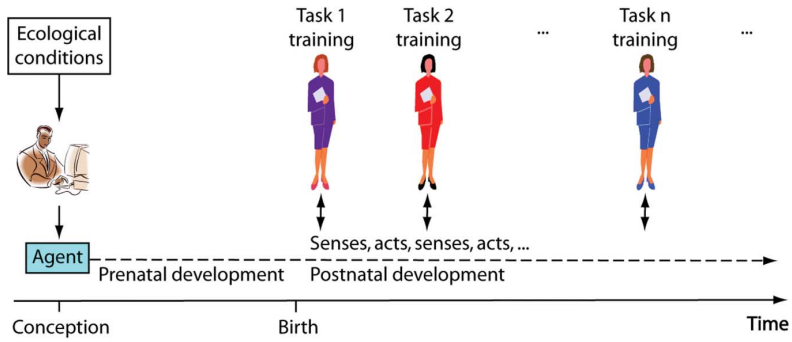

Fig. 1. Paradigm of developmental agents, inspired by human mental development. No task is given during the programming (i.e., conception) time, during which a general-purpose task-nonspecific developmental program is loaded onto the agent. Prenatal development is used for developing some initial processing pathways in the brain using spontaneous (internally generated) signals from sensors. After the birth, the agent starts to learn an open series of tasks through interactions with the physical world. The tasks that the agent learns are determined after the birth.

of knowledge. However, any mixed approach is task-specific [20]. I provide reasons why the internal representations inside the brain fully autonomously emerge and why this capability is of essential importance not only for understanding the natural brain-mind but also for computers and robots to address a series of well-known bottleneck AI problems, such as brittleness, scalability, and computational complexity. These bottleneck problems are "here and now" for any partially autonomous robot and any AI system.

3) Conceptual Comparison of Approaches: Therefore, a hallmark difference between traditional AI approaches and autonomous mental development [20] is task specificity. All the prior approaches to AI are task specific, except the developmental approach. Table I lists the major differences among existing approaches to AI and to modeling the mind. An entry marked as "avoid modeling" means that the representation is emergent from experience.

Traditionally, given a task to be executed by the machine, it is the human programmer who understands the task and, based on his understanding, designs a task-specific representation. Depending on different approaches, different techniques are used to produce the mapping from sensory inputs to effector (motor) outputs. The techniques used range from direct programming (knowledge-based approach), to learning the parameters (learning-based approach), to handcrafting sensorimotor rules for resolving behavior conflicts (behavior-based approach), to genetic search (genetic approach). Although genetic search is a general-purpose method, the chromosome representations used in artificial genetic search algorithms are task specific. Additionally, the cost of genetic search for the major components in the brain-part of developmental program (DP) seems intractably high for human-level performance even with full AMD.

Using the developmental approach inspired by human mental development, the tasks that the robot (or human) ends up doing are unknown during the programming time (or conception time), as illustrated in Fig. 1. The ecological conditions that the robot will operate under must be known, so that the programmer can design the body of the robot, including sensors and effectors, suited for the ecological conditions. The programmer may guess some typical tasks that the robot will learn to perform. However, world knowledge is not modeled and only a set of simple reflexes is allowed for the developmental program. During "prenatal" development, internally generated synthetic data can be used to develop the system before birth. For example, the retina may generate spontaneous signals to be used for the prenatal development of the visual pathway. At the "birth" time, the robot's power is turned on. The robot starts to interact with its environment, including its teachers, in real time. The tasks the robot learns are in the mind of its teachers. In order for the later learning to use the skills learned in early learning, a well designed sequence of educational experience is an important research issue.

Many variations of approaches in the field of AI are based on a good intention of convergence toward more human-like intelligence. They emphasize other factors such as embodiment and activeness that are also implied for any developmental agent [e.g., enactive AI [46] and various approaches to artificial general intelligence (AGI) [15] which however use symbolic representations]. They refine the learning-based approach and behavior-based approach by taking advantage of both. Other combinations of the four approaches in Table I have also been investigated. For example, George & Hawkins 2009 [16] intended to model some cerebral cortex. Because of their use of handcrafted concept contents (e.g., features) and the boundaries between concepts that are intrinsic to the symbolic Hidden Markov Models (HMM) that they use, their representations

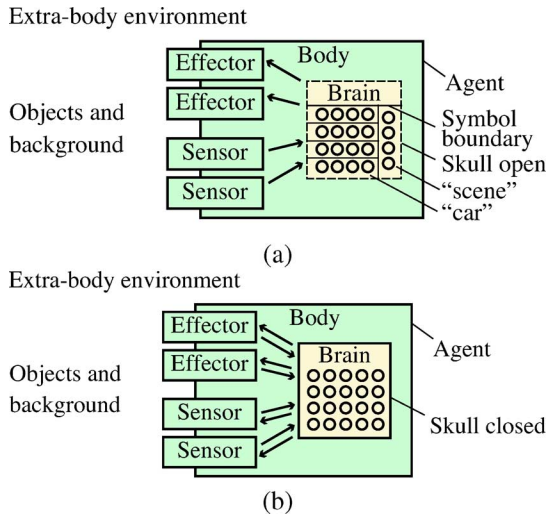

Fig. 2. Agents using (a) symbolic representation and (b) emergent representation. In (a), the skull is open during manual development—the human programmers handpicked a set of task-specific *extra-body* concepts using human-understood symbols and handcraft the boundaries that separate *extra-body* concepts. In (b), the skull is closed during autonomous development—the human programmers only design task-nonspecific mechanisms involving only *intrabody* concepts for autonomous mental development and the internal representations autonomously emerge from experience. Many neural networks do not use the feedback connections in (b).

seem to fall into the symbolic category, not emergent according to our definition later (see Fig. 2).

4) Necessity of Autonomous Development: There are a series of reasons for the necessity of autonomous mental development [20], [47], [48]. Among the top compelling ones are:

1) **high muddiness**: the muddiness (in the sense of Weng 2009 [48]) of many tasks (including task environments) is too high to be tractable by a static large-scale handcrafted representation;
2) **cost of learning**: higher mental skills (e.g., complex attention and abstract thinking) require rich learning. Then, learning must be autonomous;
3) **creativity**: complex mental skills indicated by creativity at different ages cannot be handcrafted since no human knows the contents.

The task-nonspecificity of the developmental program represents a departure from traditional understanding and modeling agent architectures. In the following two sections, I review major published agent architectures emphasizing internal representations.

III. Symbolic Representations

Symbols, including their meanings represented by languages, are created by a human society for communication among humans. Each symbol, either an actual term that describes its meaning (e.g., "cat," "young," or "kitten") or its label "A," has a well-defined meaning understandable by a group of human adults involved in the discussion. However, the agent itself does not understand the meanings of each symbol. For example, all the text strings in Fig. 8(b) are only in the minds of the human group. The machine has no clue about the meanings of any of the text strings.

Without a human language, many meanings cannot be effectively conveyed from one person to another. Therefore, the first representation type used by the AI community and the Cognitive Science community is symbolic, although this type of representation is not necessarily sufficient to characterize the human brain's internal representation and human intelligence.

A. Symbolic Representation

Since there have been several key concepts that tend to be misleading in terms of representation, we need to be specific in terms of definition for the purpose of this review. First, we define the term "internal."

Definition 2 (Internal): The term *Internal* in the "brain" of an agent refers to everything inside the brain's skull, which is not directly exposed to the external world and cannot be directly supervised by the external world.

Biologically, the term "brain" in the above definition should be considered the central nervous system. The term "brain" is used mainly for its more intuitive common sense.

The retina is an sensory port of the brain that is not internal, as the external environment can directly "supervise" the image on the retina. The motor neurons in the muscles are also not internal, as the external world (e.g., mother's hand) can supervise the motor neurons via direct physical guidance (e.g., guiding child's hand). In general, we consider that the sensory port (sensors) and the motor port (effectors) of the brain are not internal, since they are exposed to the external physical world, which includes the body and the extra-body environment, as illustrated in Fig. 2.

A symbolic (internal) representation, illustrated in Fig. 2, is defined in this review as follows:

Definition 3 (Symbolic): A symbolic representation in the brain of an agent contains a number of concept zones where the content in each zone and the boundary between zones are human *handcrafted*. Each zone, typically denoted by a *symbol*, represents a concept about the *extra-body* environment.

In order to be usable as intended, a symbolic representation has the following three characteristics indicated by the three italic words in the above definition, as shown by the example in Fig. 2(a):

1) **extra-body concepts:** the *concepts* to be represented include those about the *extra-body* environment—outside the body of the agent (e.g., oriented edges, human faces, objects, tasks, goals), in addition to those about the intra-body environment, such as body components (e.g.,muscles and neurons) and inborn reflexes (e.g., sucking);
2) **symbol:** there is an association between each computer *symbol* (e.g., ASCII code) and a text string in a natural language so that the human programmers understand the meaning of each computer symbol (e.g., text "car" in English) but the agent does not;
3) **handcraft:** the human programmers of the agent handcraft the *concept contents and boundaries* inside the "skull" of the agent, as illustrated in Fig. 2(a). By "skull," we mean the imaginary encapsulation that separates the agent "brain" from the remainder world.

Allen Newell [3], Zenon Pylyshyn [49], and Jerry Fodor [50] thought that the symbol systems popular in AI and computer science can well characterize human cognition.

Apparently, symbolic systems do not characterize sensory perception well. S. Harnad [51], [52] raised the symbol grounding problem, for which he proposed symbolic representations in which symbolic names in the category taxonomy be strung together into propositions about further category membership relations (e.g., "zebra" is represented as "horse" and "stripes") ([52], p. 343).

A major advantage of symbolic representation is that the elements of the representation have prespecified abstract meanings so that the representation is easier for humans to understand.

Such a symbolic representation follows a design *framework* that is meant to be general (e.g., ACT-R). Many expert systems can be built based on a framework. However, each design instantiation that follows such a framework is still handcrafted and static once handcrafted. Two things must be done manually in each instantiation—hand-picking and mapping. Given a problem to be solved by the machine, the human programmer must: 1) hand-pick a set of important concepts for the given task and then; 2) manually establish a mapping between each concept (e.g., location or object) and an symbolic element (e.g., a node) in the representation (e.g., for training). That is, the learning process cannot start until a task-specific representation is handcrafted and the external-to-internal mapping is manually established.

Thus, using a symbolic representation, the human programmer is in the loop of task-specific representation design and needs access to the internal representation during learning. The agent is not able to learn skills for an open number of simple-to-complex tasks through autonomous interactions in open complex environments—the hallmark of autonomous mental development. There are also some major technical issues. For example, *the curse of dimensionality* [53] is a well known problem of such a static set of hand-selected concepts (or features), e.g., adding more handcrafted features does not necessarily improve the recognition rate.

There are two types of symbolic representation, symbolic monolithic and symbolic contextual. These are new definitions, as far as I know. In the former, a monolithic data structure is used to describe the modeled part of the extra-body environment. In the latter, states are defined each of which represents a different context but not the entire modeled extra-body environment.

B. Symbolic Monolithic

The agent architectures in this category use a monolithic data structure to represent the extra-body environment that the human programmer is interested in. A hypothetical agent architecture of this type is shown in Fig. 3. The typical symbolic concepts hand-picked for the representation include location, size, color, and curvature. It is worth noting that if multiple monolithic maps are used, e.g., one for depth and one for curvature, the corresponding representation is still monolithic, because it just uses different entries to store different concepts.

Much work in AI used symbolic monolithic representations.

1) Spatial: In computer vision, the 3-D shape of an object and the 3-D motion of an object have been a major subject of study. D. Marr *et al.* were among the earlier pioneers who introduced computational modeling into understanding of human vision. Fig. 4 illustrates an example, where 3-D location and

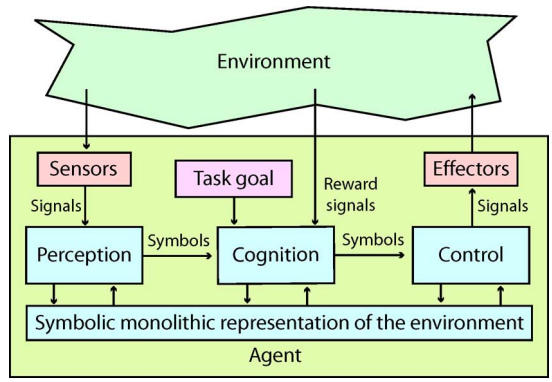

Fig. 3. Hypothetical agent architecture that uses a symbolic monolithic representation for the extra-body environment.

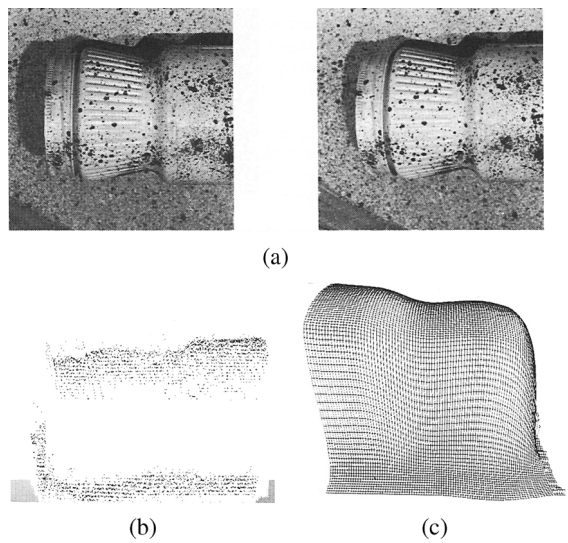

Fig. 4. Example of symbolic monolithic spatial representation: Each element in the representation corresponds to a symbolic meaning (binocular disparity) of the part of the environment being represented. (a) Two stereo images of a bottle. (b) The disparity map computed from the stereo images. (c) A smooth disparity map interpolated from (b). Adapted from Grimson [57] (a), (b), and Marr [58] (c).

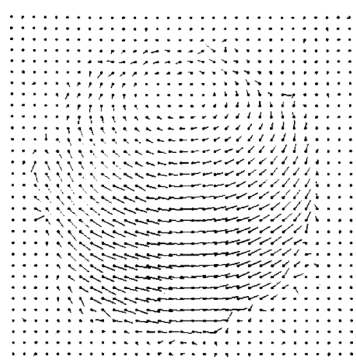

Fig. 5. Example of symbolic monolithic short-term temporal representation: Each needle in the spatial map corresponds to the motion speed of the pixel at the location. Adapted from Horn & Schunck [59].

3-D shape of the object being modeled are of primary interest. The representation is monolithic because a monolithic partial **3-D map** (called 2.5-D) is used in the presentation. A prevailing number of published methods for stereo vision used a symbolic monolithic representation (e.g., Dhond & Aggarwal [54], Weng *et al.* [55], and Zitnick & Kanade [56].)

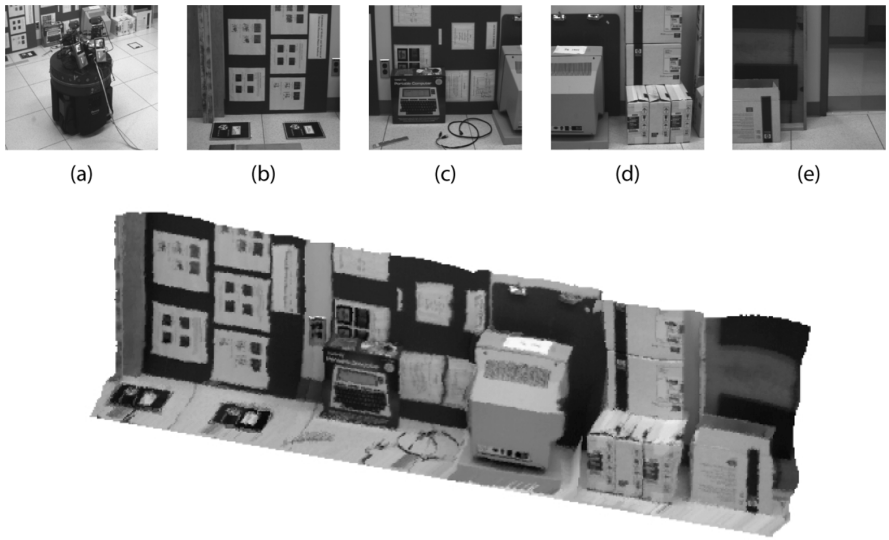

Fig. 6. SLAM by a robot from a transitory stereo image sequences with 151 time frames: (a) the robot; (b) the left image of frame 0; (c) the left image of frame 50; (d) the left image of frame 100; (e) the left image of frame 150. Lower row: The reconstructed 3-D surface optimally integrated from 151-frame transitory image sequence, shown with original intensity viewed from an arbitrary direction. Adapted from Weng *et al.* [70].

2) Temporal: The temporal information of the modeled world has been studied in the subject of motion. Algorithms have been proposed to compute the motion velocity of individual pixels, called **optical flow**. Horn & Schunck [59] were among the first to propose a method to compute the optimal flow field from a sequence of images. Fig. 5 illustrates the monolithic spatio–temporal representation used. A rich collection of motion estimation methods have been published (e.g., from optical flow [60] and from large motion disparities [61]).

Object-based motion analysis methods have three categories of assumptions: 1) a static world (e.g., [62] and [63]); 2) a single rigid object (e.g., [64]–[66]); 3) nonrigid, but of a preassumed type (e.g., elastic or articulated [67] and [68]). Human-defined features have been used for motion analysis, including intensities, points, lines, and regions.

3) Long-Range Spatio–Temporal: The goal of the symbolic monolithic representations is to model an extensive part of the environment, but each sensory view only covers a small part. Multiple sensed views have be integrated to give an extended scene map where each component in the map takes advantage of multiple observations from a moving robot.

Cheeseman & Smith [69] introduced the extended Kalman filter (**EKF**) techniques to address this problem. The relative locations of the tracked scene landmarks can be incrementally integrated by EKF through multiple views of a mobile robot and their estimated locations are provided along with estimated covariance matrices of errors. The transformation from one view to the next is handcrafted and nonlinear. Local linear approximation is used by EKF.

Weng *et al.* [70] studied *transitory sequences* which are those image sequences whose first view does not share any scene element with the last, as shown in Fig. 6. Their incremental optimization algorithm [70] generated an image-intensity mapped 3-D world map associated with estimated viewer locations and poses along the trajectory. This was done by optimally integrating a transitory image sequence using automatic pixel-based stereo matching and automatically selected dynamically tracked visual features. A view of the automatically constructed 3-D world map is shown in Fig. 6. This problem was later called simultaneous localization and mapping (**SLAM**) [71]–[73], where range sensors are often used for environment sensing instead of the stereo cameras used in [70]. An example of the FastSLAM algorithm [74] is shown in Fig. 7.

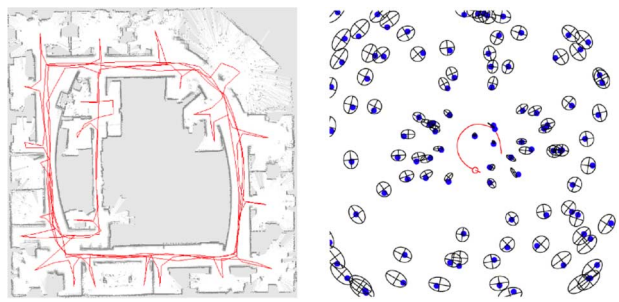

Fig. 7. Left: A 2-D range map constructed by FastSLAM from the laser scan data acquired by a mobile robot which traveled along the marked red trajectory (adapted from Hähnel *et al.* [74]). Right: Representation of errors of (simulated) landmarks as uncertainty ellipses (adapted from Montemerlo *et al.* [75]).

The theoretical analysis in Weng *et al.* [70] showed that the error of the reconstructed world map through a *transitory sequence* intrinsically suffers from at least a linearly increasing accumulation of error determined by the Cramér-Rao lower error bound. The farther the robot goes, the larger the expected error is in the estimated location of the world map. This indicates a major problem of monolithic representation.

The above systems do not assume a known type of object. Model-based map reconstruction is a very active subject in computer vision. Assuming that a scene contains only a human face, some model-based experimental systems can build an image-intensity mapped 3-D face model with minimal user interaction. For example, Liu *et al.* [76] reported a system that takes images

and video sequences of a face with a video camera. Five manual clicks on two images are needed to tell the system where the eye corners, nose top, and mouth corners are. Then, the system automatically generates a realistic looking 3-D human head model and the reconstructed model can also be animated.

4) Model-Based Recognition: Model-based recognition has a long history. Compared with previous tasks whose goal is to construct 3-D shapes and other related properties such as motion information, model-based recognition intends to provide more abstract information about the environment—object type or other related information. It is the human programmer who handcrafts a model about the type of objects to be recognized, such as human faces, human bodies, and cars. A flexible matching scheme is used to match the model to a given image, typically assuming that the object modeled is present [77]–[84]. HMM has been used [85]–[87].

5) SMPA Framework: Not all the studies reviewed here deal with cognition and robot control. A widely accepted, but wrong, notion in the field of AI is that perception is a module in an agent. This module inputs images and outputs an application specific scene description Horn ([88], p.13), Ballard & Brown ([89], p. xiii) for later processing inside the agent. It was hoped that perception problems would be resolved one day as a module. Much of the research in the robotics field has been concentrating on action generation and control (e.g., dancing or walking by humanoids) which tends to impose relatively weak demands on knowing the environment. Unfortunately, much of the agent research has drastically simplified or bypassed perception, concentrating on the cognition module. Later in the review, we will see that both perception and cognition require actions. The monolithic representation generated is suited for the sense-model-plan-act framework (criticized by Brooks [90]), or **SMPA** for short: 1) sense the environment; 2) build a model of the sensed part of the environment, using a monolithic representation.; 3) plan or replan the course of actions; 4) Act according to the plan. Although the above SMPA sequence can be updated when the robot receives new sensory information, the framework is penalized by the requirement of building a monolithic model.

6) Comments: There are some obvious advantages with symbolic monolithic representations:

1) **intuitive**: it is intuitive and easy to be understood by a human;
2) **suited for visualization**, as Figs. 4–7 show.

There is a series of disadvantages with a symbolic monolithic representation. Some major ones are:

1) **wasteful**: the representation is a model about the environment, but only a very small part of the environment is related to the action at any time. Further, if a robot needs to explore an open-ended world, no fixed-size monolithic representation will be sufficient;
2) **insufficient**: the representation is often not sufficient to generate all the desired behaviors. For example, a range map is not sufficient for a robot to read a room number;
3) **low level**: yhe representation only deals with few low-level symbolic attributes, such as location, distance, and size. These attributes are not abstract enough for action generation.

Is it true that the cerebral cortex uses symbolic representations, in the sense that its neurons represent orientation, direction of motion, and binocular disparity? The discussion in Section IV argues that the partition of the responsibilities among neurons is not based on such human imagined symbolic meanings.

C. Symbolic Contextual

In a symbolic contextual representation, only information that is related to the immediate actions is represented as an abstract symbolic state. The state is a representation of many equivalent contexts. Therefore, in general, symbolic contextual representations are more purposive than symbolic monolithic representations.

Yiannis Aloimonos [91] pointed out correctly that computer vision needs to be purposive in the sense that a monolithic representation is not always necessary. However, symbolic contextual representations are not sufficiently powerful for dealing with purposive vision in general open settings. Such a representation has been often used for modeling a symbolic or simplified microworld. Examples of such microworlds are: a given word in speech recognition, a given simplified human body action in human action recognition, a given manipulatory action in robot manipulation.

Symbolic contextual models have used finite automata (FAs) [92] as the basis framework. Examples include **ACT-R** [93], **Soar** [94]–[96], **Neuro-Soar** by Cho *et al.* [95], **CYC** [97], **Bayesian Nets** (also called **Semantic Nets**, **Belief Nets**, **Bayesian parsing graph**, Graphic Models) [37], [19], [98]–[101], [43], **Markov Decision Processes** [102], [103], the partially observable Markov decision process (**POMDP**) [104], [102], the sensory classification counterpart of POMDP—**HMM** [103], [105], {106], [86], [101], the **Q-learning Nets** [107], and other **reinforcement learning nets** [108]–[113].

We should start with **finite automata** since they are the most basic among all symbolic contextual architectures.

1) FAs: A *deterministic* finite automaton FA for a microworld consists of m symbolic states, $Z = \{z_1, z_2 \dots z_m\}$, as indicated in Fig. 8. In contrast to the continuous monolithic representations in Section III.B, the set Z of states typically consists of only a finite number of discrete symbolic contexts that the system designer concentrates on as samples of context space.

The inputs to the FA are also symbolic. The input space is denoted as $X = \{x_1, x_2 \dots x_l\}$, which can be a discretized version of a continuous space of input. In sentence recognition, e.g., the FA reads one word at a time. l is equal to the number of all possible words—the size of the vocabulary.

For example, reading input sentence "I already agreed," z_1, z_2, and z_3 correspond to three subsequences "I," "I already," and "I already agreed," respectively, that the FA needs to detect. Fig. 8 gives two examples. It is important to note that the "meanings" of the input labels and the text-described

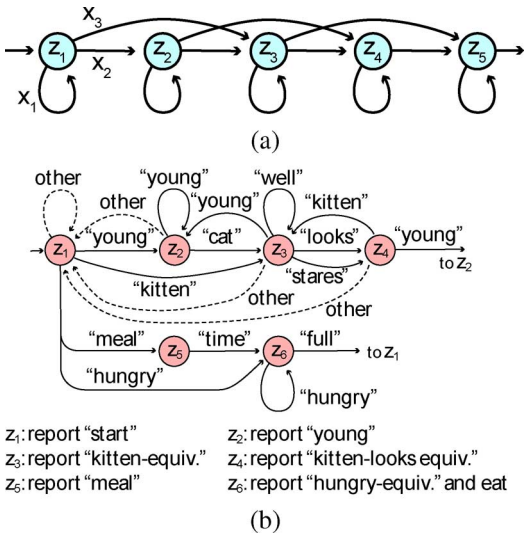

Fig. 8. Simplest symbolic contextual representations: Finite Automata (FAs). (a) A left-right FA in which no state can transit to a state to its left, commonly used in speech recognition. Each circle indicates a symbolic context, since the network represents a symbolic context (e.g., a hand-selected word). (b) A more general FA. It starts from z_1. A label "other" means any symbol other than the symbols marked from the state. The "other" transitions from the lower part are omitted for clarity. All the English input labels and the lower three rows of English text are only meant to facilitate human-to-human communications, not part of the FA and not something that the FA understands.

meanings of each state z_i in the lower three text rows of Fig. 8(b) are *only* in the mind of the human designers to facilitate human-to-human communications. The FA does not understand such meanings. This is a fundamental limitation of all symbolic contextual representations.

A regular FA can be extended to an *agent FA* [114] which, at each time, inputs a symbol $\mathbf{x}_i \in X$ and outputs its state $z \in Z$. In each state, cognition is part of action (e.g., vocal action or manipulation).

For an FA with m states, l inputs, the set of transitions is $\{(z, x, z') \,|\, x \in X, z \in Z, z' \in Z\}$. The number of rules to be designed is lm, since given any state z and any input x, the next state z' is uniquely specified.

If the inputs are all correct, the architecture FA works well. Good examples are numerous, from a word processor, to a graphic user interface, to a video game.

2) Hierarchical FAs: Often, the symbolic concepts designed by the human designer are not flat and instead, they have a hierarchical structure. For example, in sentence recognition, words are at a lower level while sentences are at a higher level. Also, the detector of a sentence needs to detect consecutive words. Fig. 9 illustrates the architecture of a hierarchical FA (**HFA**). States are grouped to form higher states. The global state of a three-level hierarchical FA is represented by a 3-tuple (y_{1i}, y_{2j}, y_{3k}), where y_{li} represents the ith local state at level l. It is often that the same lower level FA is used by multiple higher states. For example, in speech recognition, two sentences may share the same word detected by the same FA at the lower level. The arrangement of such a shared word is also handcrafted.

Many of the published cognitive representations and the associated cognitive architectures belong to this category. Additional structures are specified to model other information

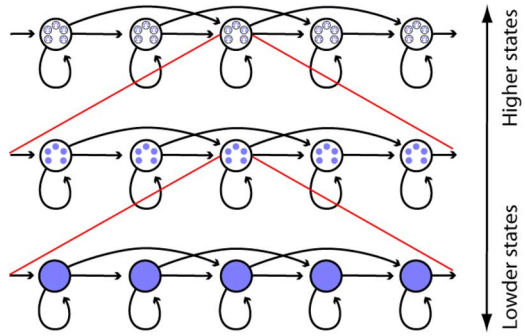

Fig. 9. Architecture of hierarchical FAs. Every higher state triggers an FA at the next lower level whose first state is the initial state and the last is the goal state. For clarity of the drawing, only the state at the middle has its triggered FA shown.

used by the agent. The following are a few samples of such architectures:

PSG: PSG seems the first production system that had an implemented computer program. Newell [2] proposed that his production system PSG models the mind. A PSG consists of symbolic production rules as if-then or condition-action pairs. The condition is incorporated by the context state and inputs. The action is represented by the state output.

ACT-R: ACT-R is a family of cognitive architectures, developed by Anderson *et al.* [93], [115], aiming at modeling human behaviors. It has different modules, each processing a different type of information, such as sensory processing, intentions for goals, declarative module for declarative knowledge, and action module for procedural knowledge. The units in which knowledge is represented in working memory are defined as chunks. The goals of the task are represented by utility, which is the difference between the expected benefit and the expected cost. Learning in ACT-R can change its structure (e.g., a constant becomes variables with new substructures) and statistical parameters (e.g., expected cost).

Soar: Soar is another family of cognitive architectures [94], [96] based on Production Systems. Procedural knowledge is represented as production rules that are organized as context-dependent operators which modify internal states and generate external actions, as illustrated in Fig. 10. A later version [116] of Soar added episodic memory and semantic memory. The former encodes a history of previous states, and the latter contains declarative knowledge. Unlike ACT-R, the task and subtasks are formulated as attempts to achieve goals and subgoals. Thus, when knowledge is not sufficient to select an operator for reaching a goal, an impasse occurs, during which Soar allows the teacher to select a new subgoal or specify how to implement the operator. Different learning mechanisms are used in Soar for learning different types of knowledge: chunking and reinforcement learning for procedural knowledge and episodic and semantic learning for declarative knowledge.

ICARUS: Proposed by Langley *et al.* [117], ICARUS has two types of knowledge, environmental concepts and skills for achieving goals. Both types are hierarchical. Environmental symbolic objects are placed into a perceptual buffer for primitive concept matching. Matched instances are added to short-term beliefs as context, which triggers the matching for higher-level concepts. In this sense, ICARUS models the

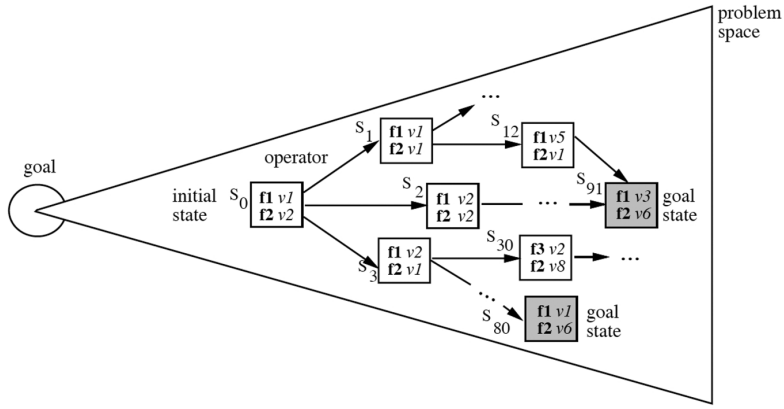

Fig. 10. Instance of Soar is a solver for a given task (problem). Squares represent states. A shaded goal represents a goal state of the subtask. Inputs are represented as features (e.g., $\mathbf{f}_1$, $\mathbf{f}_2$), with values (e.g., v_1, v_6) as condition of state transition. Adapted from Laird *et al.* [94].

sequence of matching over the hierarchical structure of parts of the world objects.

3) Probabilistic Models: Randomness is due to a lack of knowledge (uncertainty) or a simplification in modeling. For example, if all the detailed causality of tossing a coin is known and accounted for, there is no randomness with coin tossing. This perspective is useful for us to understand emergent representations, which avoid manually modeling the environment.

Tools of mathematical probability and statistics enable human modelers to improve the performance of agents without modeling details that are difficult to model correctly. If the state is deterministic (completely observable) but transition to the next state is probabilistic (e.g., due to input uncertainty or action uncertainty), the corresponding model is called a **Markov chain**. If the state that the system is in at any time cannot be determined correctly, we call that the state is partially observable or the state is hidden.

Depending on whether one needs a one-shot decision or sequential decisions, two types of Markov models have been used, **HMM** [105], [118], [119] and Markov decision processes (**MDPs**) [104], [120], [102], respectively. By sequential decisions, we mean that the final outcome depends on a series of actions while each subsequent environment depends on previous actions (e.g., chess playing and navigation).

a) One Shot Decision: If HMMs are used for classification of temporal sequences, a different Markov model is used for detection of a different temporal sequence as shown in Fig. 11. Researchers on speech recognition [103], [121], [122] and computer vision [123], [124], [98], [86] have used HMM for classifying temporal sensory inputs. HTM by George & Hawkins [16] is a handcrafted, HMM based symbolic model, although it was inspired by biological cortical computation.

To classify n sequences (e.g., n spoken words), n HMMs are needed, one for each word. For a more sophisticated system, more than n HMMs are used so that a smaller within-class variation is dealt with by each model. For example, different HMMs are used for male speakers and female speakers.

The internal representation of each HMM here is **subsymbolic**—representing finer features within a hand-selected symbol. A representation containing subsymbolic components is still symbolic according to the Definition 3. In speech recognition, e.g., the meaning of each node in HMM is typically not explicitly determined. Instead this is done through a preprocessing technique (e.g., k-mean clustering for a given word) before HMM parameter refinement (e.g., using the Baum-Welch algorithm). This is because a local refinement algorithm, such as the Baum-Welch algorithm, only update the probabilities of the transitions among defined symbolic states, but does not define the symbolic states. The meaning of each HMM is hand-defined (e.g., represent a word "Tom") and thus the meaning of each node in the HMM is still handcrafted. This practice of assigning a different HMM to a different word is not possible for autonomous development, since: 1) the tasks that the agent will learn are unknown to the programmer before the agent's "birth;" and 2) the internal representation of the agent is not accessible to the human programmer after the agent's "birth" [20]. For example, we cannot assign a cortical area to deal with a particular word for the above two reasons.

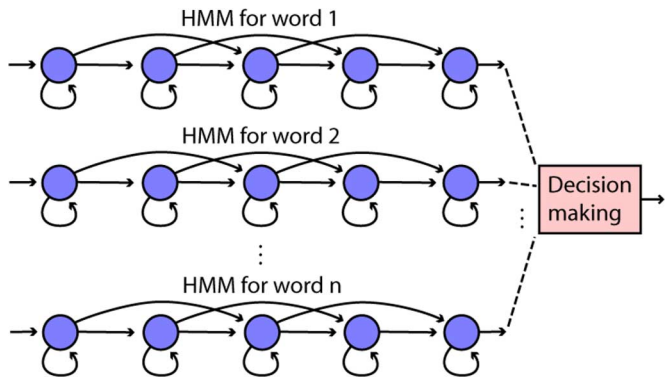

Fig. 11. Architecture of an agent that uses HMMs for classification of dynamic temporal sequences. A different HMM is used for modeling a particular type of temporal sequence. Taking chunks from a temporal sequence, all n models compute their probabilities in parallel. Each model needs to update the probabilities of all its states for every chunk of input sequence. To deal with the time warping of the input stream, each model estimates the best time sequence of its state transitions (e.g., different amounts of time to stay in each state), so that the highest probability is computed from all possible time warping for the word. The module of decision making reports the classification label from the HMM that gives the highest probability. During training, each HMM only takes samples of the assigned word. Therefore, manual internal access to the network internal representation is necessary.

b) Sequential Decision: This class includes all probabilistic models that are based on symbolic states, including

Bayesian nets, belief nets, semantic nets, and the Markov decision process.

If the model is such that each state may issue an action, the corresponding probabilistic model is called MDP. The decisions from a MDP is sequential in the following sense: A task is accomplished by a sequence of actions, issued from the sequence of states. For example, the task of moving from spot A to spot B is accomplished by a sequence of local motions executed by a robot.

The partially observable MDP (**POMDP**) has been used for range-based robot navigation [125], [126]. In contrast with the HMM-based systems which model sensory temporal context, an POMDP models the entire environment. Given each observation chunk, the model needs to update the probability of every state. At each time step, the state with highest probability can be considered the current state and the corresponding action is issued.

4) Behavior-Based Methods: The **subsumption** architecture proposed by Brooks 1986 [127], [128] belongs to the class of symbolic contextual representations. However, it emphasizes decomposition of behaviors, instead of context. Higher level behaviors subsume lower-level behaviors. For example, collision avoidance is a lower level behavior, and path planning is a higher level behavior. **Kismet** and **Cog** [129], [130] are more extensive versions of the behavior based architecture. The behavior-based method has clear advantages over symbolic monolithic representations. However, coordination of different behaviors requires handcrafting which is subject to failures for complex tasks.

5) Object–Action Framework: Ballard & Brown [131] argued that Gibson's precomputational theory eschewed any explicit representation. They proposed that cooperative sensorimotor behaviors can reduce the need for explicit representation. Aloinomos [91] proposed that vision should be studied in the context of purpose and actions. Only part of the visual world that is immediately related to the purpose and the current action (e.g., reaching) needs to be processed. This is exactly the idea of contextual representation. This line of new thinking has been adopted by an increasing number of robotic vision researchers. Wörgötter *et al.* [132] reported their work under a framework of object action complexes, which correctly emphasizes the conjunction between object (sensory) and action (motor), but using symbolic representations. Cohn *et al.* [133], Vernon [134] and many others in computer vision correctly stressed the use embodied actions as an integrated part of visual processing for objects, while using symbolic representations.

6) Learning in Symbolic Systems: Many symbolic systems include mechanisms of learning, since manually specifying all the agent structure detail and parameters is intractable for large problems.

MDP, HMM, and POMDP have well formulated learning algorithms. The objective of their learning is to compute the values of the designed probability parameters. Online learning algorithms update the probability values at every time instant. An agent with m states potentially requires an update of m^2 transition probability values at each time step. Therefore, with m states for each HMM, time length t and w HMMs, the time complexity for classification using a learned system is on the order of wtm^2 [105]. When m is exponential in the number of concepts (see Section V-F), this scheme becomes impractical for large problems.

Another more efficient approach to learning is to avoid using probabilities. Q-learning [107] uses a reward value r from the environment to estimate the value of each action at each context state. It assumes that the state is observable and at each state there is a finite number n of actions. This online algorithm updates the estimated value $Q(s, a)$ of action a at state s using a $O(n)$ time. However, the convergence of $Q(s, a)$ by this algorithm requires that the number of visits of every possible state-action pair approaches infinity.

There are many more cognitive models whose goal is to model the mind or even the brain, but their modelers used symbolic representations. Examples include CLARION by Ron Sun [41], HTM by George & Hawkins [16], GNOSIS by John Taylor *et al.* [135], ARTSCAN by Stephen Grossberg *et al.* [136], Albus [42], and Tenenbaum *et al.* [43].

The above examples reviewed indicate an existing approach to mental architecture research—extra-body concepts are handcrafted into the representation. How many extra-body concepts to be handpicked and to be modeled is a human design choice.

However, several major factors suggest that symbolic modeling is inappropriate for large problems and, thus, also insufficient for understanding the brain. In the following, the first two problems are from my personal views. The later two problems are well known [19], [99].

1) **The state-size problem**: An exponential number of states is needed for complex tasks, as we discuss below. For example, although a chess game has a finite number of states, the number of all possible board configurations is an exponential function of the number of pieces on the board. The number of states may be larger than the storage size of any practical computers. Therefore, there seems no practical way to specify all possible state rules. Here comes a basic problem of symbolic modeling: partition all the possible task states into a much smaller, tractable number of symbolic states. As a smaller set of rules is used for many task states, the output does not fit all these original states. This causes the well known *brittleness problem* of a symbolic system: As soon as the task situation does not fit the symbolic model, the system breaks. Worse, the system fails miserably as soon as the system makes a mistake, since a wrong state leads to further wrong actions and wrong states. Adding probability to a symbolic model only alleviates the brittleness, but is not always sufficient to recover from such representation failures.
2) **The attention problem**: The input space X corresponds to sensors, and thus it is a high dimensional continuous space (e.g., the dimension equals to the number of pixels in an image). At a different state, different subparts of the input space X are related to the next state. The locations and shapes of these subparts are unknown and dynamic (attention problem). These subparts are also related to system state in a highly nonlinear, complex, and dynamic way. The total number of such subparts is an exponential function of the sensory dimension. No human designed symbols are

sufficient to deal with such an exponential number of subparts.

3) **The frame problem**: This is a problem of expressing a dynamic world in logic by a disembodied agent without explicitly specifying which environmental conditions are not affected by an action. Some facts remain the same, but others may change. In general, this problem is related to the maintenance problem of a disembodied knowledge base.
4) **The ramification problem**: Actions of an agent may fail and cause side effects that have not been modeled by a symbolic representation. For example, an object may not be successfully picked up by a robot hand, the robot wheels may slip, and small items on an object may fall when the object is picked up. The ramification problem is due to the disembodied manual representation design—the agent is not able to self-generate internal representation.

7) Comments:

a) Symbolic Versus Emergent: It is important to note a fundamental difference between a symbolic representation here and the emergent representation to be discussed later. In a symbolic representation, a symbol corresponds to an atomic concept describing a property of the environment. One symbol is sufficient and is unique within the set of handcrafted concepts. In an emergent representation, no internal element corresponds to any concept that can be verbally communicated among humans.

b) Probabilistic Is Not Distributed: The flat system model in Fig. 8 and the hierarchical model in Fig. 9 still use a symbolic representation because the underlying fact still has one active state at any time, not multiple.

c) Machine's Ignorance of Internal Symbols: The meaning of a symbolic representation, either monolithic or contextual, is only in the mind of the human designer, not sensed by the agent. Therefore, the agent is not able to understand the meaning of the representation. For example, as the agent does not "know" that an internal HMM M_1 is for detecting word w_1, it does not understand the word w_1 and cannot change the internal algorithm to improve the recognition of w_1.

There are some advantages with symbolic contextual representations:

1) **intuitive**: the design of the representation is in terms of human well understood symbols. Thus, the representation is easier to comprehend;
2) **unique**: a symbol is unique. There is no ambiguity to tell whether two symbols are the same.

There are also a series of disadvantages with a symbolic contextual representation. Some major ones are listed below.

1) **Brittle**. The more symbolic concepts and symbolic models are manually designed, the more rigid walls are built among them. Generalization across the walls is disallowed by design. The automatic model applicability checker—at any time which model applies and others do not—is an unsolved hard problem [137], resulting in brittle systems in real-world settings.
2) **Wasteful**. Although this representation is more efficient than symbolic monolithic representation, it does not use brain-like emergent representation with which the state of the brain is determined by the corresponding firing pattern. In Section V, we will see that a moderate number of concepts require a large number of states, larger than the number of neurons in the brain.
3) **Insufficient**. A symbolic representation does not provide sufficient information to allow the use of relationships among different symbolic states. In contrast, using emergent representations, different firing patterns naturally allow a determination of the goodness of match through the neuronal responses of the next layer.
4) **Static**. A machine using a symbolic representation (e.g., Bayesian net) is not able to go beyond all possible combinations of the handpicked symbolic concepts. For example, the chess-playing program IBM Deep Blue [138] is not able to "step back" to view what it does in a larger scope and to improve its method. It does not even "know" that it does chess-playing. The same is true for the Jeopardy Game program IBM Watson.
5) **Nondevelopmental** Although a symbolic representation (e.g., Bayesian net) allows learning, its learning is not fully autonomous (not AMD). The human teacher selects a learning type (e.g., classical conditioning, instrumental conditioning, fact learning, procedural learning, or transfer learning). He feeds information into a particular module in the "brain" during training. The "skull" of the machine "brain" is not closed and the "brain" inside the "skull" is not autonomous.

It appears that symbolic representation might have a hard ceiling in bottleneck problems of AI, including vision, audition and language understanding.

IV. Emergent Representations

The brain and artificial neural networks share the same basic characteristic: the internal representation is emergent from learning experience, regulated by the genes or a human designed learning program [34], [35], [139], [20], [140], [28].

According to the Merriam-Webser online dictionary, "symbol" is an arbitrary or conventional sign used in writing or printing.

The meaning of a "symbol" in human communication is slightly different from what is meant in computer programming.

A. Computer Symbols

Definition 4 (Computer Symbols): A system of computer symbols is a basic coding (e.g., ASCII coding) for different sensations and actions of symbols used in human communications. Such a coding must be unique for each class of equivalent sensations and actions among humans who communicate.

For example, regardless how different the lower case "b" appears on a computer screen or on a printed page, the computer industry agreed that its ASCII code in decimal form is 98. In doing so, the computer industry assumes that a normal human adult has no problem to correctly relate every his sensation of "b" to the unique ASCII code 98. That is, computer symbols are "consensual" among human adults. The uniqueness of the code is important since different parts of a computer program and different computer programs can all communicate correctly without ambiguity.

However, it appears that the brain never directly receives a computer symbol (e.g., ASCII code) nor directly produces directly a computer symbol as we discussed below.

B. Symbols in Human Communications

The brain receives raw signals from eye, ears, and touch. For example, the printed text "b" is read as different instances of nonsymbolic video on the retina. Therefore, the brain does not seem to receive any computer symbol directly since the uniqueness in the raw sensory signals is almost never true.

Every instance of an action (muscle contraction) that describes the same meaning (e.g., pronounce the letter "b") is at least slightly different. Therefore, the brain does not seem to directly produce a computer symbol either since the uniqueness in the raw muscle signals is almost never true.

A correct interpretation of a sensation by the human A is not confirmed by another human B until the corresponding action from A, often through a human language, is considered correct by the human B. For example, the human B asks: " What is this?" The human A replies: "This is a lower case 'b.'"

C. Emergent Representation

From the above discussion, it seem unlikely that the human brain internally uses a system like computer symbols, since neither the uniqueness of the sensations into the brain nor the uniqueness of the actions from the brain is guaranteed. Put in short, the brain's internal representations seem not symbolic. This seems a natural consequence from autonomous development since the skull of the agent "brain" is closed throughout the lifetime—after the agent "birth" the human programmer is not available [20] for direct supervision of the brain's internal organization.

This line of thought motivates our following definition of emergent representation:

Definition 5 (Emergent Representation): An emergent representation emerges autonomously from system's interactions with the *external* world and the *internal world* via its sensors and its effectors without using the handcrafted (or gene-specified) content or the handcrafted boundaries for concepts about the extra-body environments.

Based on Definition 3, a representation that partially uses symbolic representation still belongs to symbolic representation categorically, even if, e.g., some parts of it use neural networks.

Note that it is fine for a concept about the intrabody environments (inside the body or inside brain) to be programmer-handcrafted or gene-specified (e.g., eyes, muscles, neurons, and dopamine), because the brain has direct access to objects in such an intrabody environment during development. Innate behaviors (e.g., sucking and rooting) are intrabody concepts (e.g., receptors, neuronal connections, and muscles) but not extra-body concepts (e.g., not nipple, since a stick can also elicit the sucking inborn behavior). The genome uses morphogens (molecules locally released from cells) inside the body and the brain to guide the migration of other cells to form brain laminar layers, body tissues, and body organs [141], [142]. They also account for inborn reflexes (e.g., sucking), but inborn reflexes are not extra-body concepts.

Components of emergent representation in the brain include: neuronal responses, synaptic conductance (weight vectors), neuronal connections, neuronal plasticities, neuronal ages, cell types, and neurotransmitter types.

Human teachers can interact with the brain's sensory ports and the motor ports to teach the brain. Only in simplified cases, may a motor neuron (not part of internal representation) be supervised to represent a unique meaning, but in general each motor neuron (e.g., a muscle in the vocal tract) is used for different meanings (e.g., pronouncing different words).

For the brain to use an emergent representation, the brain's internal network must autonomously self-organize its internal representation (inside the skull) through experience. In particular, the programmer does not hand-select extra-body concepts (e.g., objects, edges). Without extra-body concepts to start with, of course the programmer does not handcraft any *concept boundaries* inside the "skull" for the agent brain. In other words, the internal self-organization is autonomous.

One may say that such restrictions will make it much harder to program an agent. Yes, this is true in terms of the knowledge required to program a developmental agent. However, these restrictions are meant for the full "autonomy" in internal brain-like mental development so that the agent can autonomously scale up (develop) its mental skills without requiring too much tedious human labor. In other words, a developmental agent is harder for a human programmer to understand, but should be an easier way for a machine to approach human level performance than a nondevelopmental agent.

D. Networks That Use Emergent Representations

An emergent representation is harder for humans to understand, as it is *emergent* in the sense that a particular meaning is not uniquely represented by the status of a single neuron or by a unique firing pattern of a set of neurons. This is in sharp contrast with a symbolic representation, where each single element (symbol) uniquely represents a particular meaning.

Recurrent connectionist models draw inspiration from biological emergent representations. The examples include McClelland *et al.* [143], Weng *et al.* [30], Elman *et al.* [34], Sprons *et al.* [144], Wiemer [145], Roelfsema & VanOoyen [146], Sit & Miikkulainen [147], Hinton [148], and Weng *et al.* [149] where the representations are mostly emergent although some of these models do not exactly satisfy our strict Definition 5 for a fully emergent representation. The epigenetic development (ED) network [114] (later called DN) further suggests that the meaning of any neuron or any brain area cannot be precisely described in a human language. Although the brain's internal representation is related to some symbolic meanings (e.g., the direction of edge features), the learning mechanisms are not rooted in such symbolic meanings.

An emergent representation is typically distributed. By distributed, we mean that response values from many neurons are necessary to complete a representation. In some special cases, the distributed representation in a cortical area is degenerated into a single neuron firing. However, this is a special case, instead of the norm of emergent representation. Firing neurons and not firing neurons all give information for the firing pattern.

E. Biological Considerations

A classical study by Blakemore & Cooper [150] reported that if kittens were raised in an environment with only vertical edges, only neurons that respond to vertical or nearly vertical edges were found in the primary visual cortex. Recently, such knowledge has been greatly enriched. Experimental studies have shown how the cortex develops through input-driven self-organization, in a dynamic equilibrium with internal and external inputs (e.g., Merzenich *et al.* [151], [152]; Callaway & Katz [153]; Gilbert & Wiesel [154]; Lowel & Singer [155]). Such dynamic development and adaptation occurs from the prenatal stage (e.g., Feller *et al.* [156], Meister *et al.* [157]) and continues throughout infancy, childhood, and adulthood (e.g., Wang & Merzenich [158], Drafoi & Sur [159]). Hosoya & Meister [160] reported that even retinal representation is adaptive. The spatio–temporal receptive fields and the response of retinal ganglion cells change after a few seconds in a new environment. The changes are adaptive, in that the new receptive field improves predictive coding under the new image statistics. However, the computational principles underlying the development (adaptation) of retinal, LGN, and cortical neurons are elusive.

It is important to note that feature maps described in the neuroscience literature were generated using human selected specific stimuli. For example, an orientation map of the V1 area [161] is generated using oriented gratings. It does not mean that the purpose of a neuron in V1 is only for detecting a particular orientation. Nor does it mean that a neuron's peaked sensitivity to a particular orientation of gratings is genetically fully specified. All feature maps seem activity-dependent [159], [28].

F. Networks Using Built-In Invariance

There have been many computational studies with a goal of building adaptive networks for pattern recognition, regression and other applications. Some networks have built-in (programmed-in) invariance. Locational invariance has been commonly designed as a built-in invariance.

Neocognitron by Fukushima [162] is a self-organizing multilayered neural network for pattern recognition unaffected by *shift in location*. **Cresceptron** by Weng *et al.* [32] has an architectural framework similar to Fukushima's Neocognitron but the neural layers in Cresceptron are dynamically generated from sensing experience. Thus, the circuits of Cresceptron is a *function of sensory signals*, but the same is not true with Neocognitron. The above two networks have built-in shift-invariance in that weights are copied across neurons centered at different retinal locations. However, they do not provide effective mechanisms for learning other types of invariance, such as size, orientation, and lighting. Inspired by biological networks, the ED network discussed below is not constrained by any practical type of invariance and it has the potential to learn any type of invariance sufficiently observed from experience.

Built-in invariance for an extra-body concept (e.g., location) seems not desirable for AMD since the concept could be necessary for some future tasks.

G. Networks Using Unsupervised Learning

The required invariance is learned object-by-object at the last stage (e.g., a classifier). The self-organizing maps (**SOM**) proposed by Teuvo Kohonen and many unsupervised variants [163] belong to this category. Kohonen [164] seems to prefer to use unsupervised SOM followed by the Learning Vector Quantization (**LVQ**) method, which is basically a nearest neighbor classifier. The self-organizing hierarchical mapping by Zhang & Weng [165] was motivated by *representation completeness* using incremental principle component analysis (**PCA**) and showed that the neural coding can reconstruct the original signal to a large degree. Miikkulainen *et al.* [166] developed a multilayer network **LISSOM** with nearby *excitatory interactions* surrounded by *inhibitory interactions*. independent component analysis (**ICA**) [167]–[170] has been propose for feature extraction. There have been many other studies on computational modeling of retinotopic networks[1] (e.g., [171], [172], [173], [174], and [175])

Some cognitive scientists believed that there exist two types of learning in the brain, explicit (conscious, action-based) and implicit (unconscious, not action-based, using episodic memory) [176], [177]. According to the neuroanatomy, hardly any area in the central nervous system does not have descending connections from motor areas. Namely, unsupervised learning seems rare among brain areas. For example, although recognition of an object does not necessarily involve an explicit arm reaching action, one of its actions corresponds to the *vocally* naming the object. For this reason, the lobe component analysis (**LCA**) [178] was proposed as an optimal model for a cortical area to extract features, but in general it takes both ascending sources and descending sources as parallel inputs.

H. Networks Using Supervised Learning

With supervised learning, the label of the class to which each sample belongs is available. The resulting features tend to be discriminative, e.g., relatively sensitive to between-class variation and relatively insensitive to within class variation. Different methods of learning give very different performances.

1) Error-Backpropagation: Gradient search has been widely used for network learning [see, e.g., a overview by Werbos [179] and LeCun *et al.* [180], cascade Correlation learning architecture (CCLA) [181] and [182]]. An advantage of the error back propagation is that the method is simple and can be incremental. A major limitation is that the error signal is not directly available from the animal's muscles (motor neurons).

Computationally, error back-propagation does not have an effective mechanism for selecting and maintaining long-term memory. Error gradient typically erases long-term memory that is necessary for other examplars. CCLA adds a new node by freezing the current network, which does not allow reuse of memory resource. There is no clear evidence that the brain uses error back-propagation.

2) Motor Output Is Directly Used for Learning: The ways to use output fall into three types.

a) Discrete Outputs: Mathematically, discrete outputs can be used as class labels to supervise feature extraction. This type

[1]Each neuron in the network corresponds to a location in the receptor surface.

seems not suited for biological networks because biology does not appear to produce discrete labels. The linear discriminant analysis (**LDA**) ([183] [184] [185], [186]) is an example of supervised linear feature extraction. The k-nearest neighbor classifier is used in the LDA feature space. The support vector machines (**SVM**) [187]–[189] have been widely used for classification tasks.

b) Continuous Outputs Form Clusters as Discrete Labels: This type uses **continuous outputs** to locally supervise ascending feature clusters. The hierarchical discriminant regression (HDR) by Hwang & Weng [190] and the incremental HDR (IHDR) by Weng & Hwang [191] use clusters in the high-dimensional output space Z as virtual labels to supervise clusters in ascending X space. The HDR engine has been used for a variety of applications, from robot visual navigation [191], speech recognition [192], skill transfer [193], to visuo–auditory joint learning [194]. In these systems, the numerical output vector $\mathbf{z}$ and the input vector $\mathbf{x}$ were combined as an expanded input $(\mathbf{x}, \mathbf{z})$ to the regressor.

c) Back-Project Continuous Outputs: ARTMAP [195] is a nearest neighbor like classifier for each bottom-up vector input in which each component is a feature value (symbolic, not natural image). After memorizing all sufficiently different ascending input vectors as prototypes, it uses descending signals to successively suppress the current prototype under consideration in order for the next prototype to be examined for the degree of match with the input. Thus, the top-down signal is not part of the features in ARTMAP, but as attention signal in the consecutive nearest neighbor search. Roelfsema & van Ooyen [146] proposed the attention-gated reinforcement learning (**AGREL**) network, which uses gradient to back-project descending signals as attention. In LISSOM and MILN, neurons take input from ascending, lateral, and descending connections. Thus, in these two systems, the top-down vector and the bottom-up vector are both part of the features to be detected, which raised new challenges in analysis and understanding. Sit & Miikkulainen [147] explained that in a recent version of **LISSOM** that uses descending connections, a neuron responding to an edge can receive descending feedback from neurons that detect corners in the next layer. In **MILN**, descending connections were shown to generate soft invariance from sensory input to motor output (Weng & Luciw [196]), to improve the "purity" of neurons (Weng *et al.* [197]), to increase the recognition rate and reduce neuronal entropy (Weng *et al.* [198]), to group neurons for similar outputs together (Luciw & Weng [199]), and to generate temporally equivalent states (Weng *et al.* [200]). Motivated by the brain neuroanatomy, Weng [114] proposed that each cortical area uses LCA like mechanisms to develop feature clusters in the parallel space of ascending input (e.g., image) and descending input (e.g., motor). The Where–What Network 3 by Luciw & Weng [201] hints how the brain learns concepts from motor supervision and uses its emergent concepts as dynamic goals on the fly to attend part of the external environment against complex backgrounds.

Among the above three types, the third type c) seems relatively closer to what the cortex uses, but this type is also the most difficult type to analyze and understand. Are such networks of general purpose, especially with regard to what symbolic models can do through handcrafting (e.g., temporal reasoning)? Recently, a positive answer to this question was proposed [114]. To understand what it means, we need to first look into the brain's need to process temporal information.

I. Networks Using Emergent Representation for Time

Different actions from the mind depend on different subsets of the past, of different temporal lengths and of different combinations of past experience. The *state* inside an agent has been used to represent all (often infinitely many) equivalent temporal contexts. In a symbolic temporal model, such states are handcrafted (e.g., HMM and POMDP). For a larger problem, the human designer handcrafts all allowable state transitions (e.g., left-to-right models or a sparse diagram for allowable state-transitions).

To generate emergent representations, the brain (natural or artificial) must self-generate states, not imposing their meanings and not specifying the boundaries of state meanings. Additionally, it is desirable not to provide a static diagram for allowable state-transitions. These goals are still not well recognized and accepted in the AI community.

Local recurrence in the network has been a common technique to generate temporal states. The **Hopfield network** [202] has a single layer, where every neuron in the layer sends its output to all other nodes except itself. The **Elman Network** and the **Jordan Network** [203] use local recurrence for a layer in a multilayer network. The **Boltzman machine** [204] is the stochastic and generative counterpart of Hopfield networks with symmetric connection weights. The Long Short Term Memory **LSTM** [39] adds a specially designed type of network modules called "error carousels" into a feedforward network so that the required *short memory* inside the network can be delayed as long as externally controlled. A Deep Belief Net [205] is a cascade of several Boltzman machines with tied weights across different Boltzman machines. Other major recurrent networks include Omlin & Giles [206], Wiemer [145], Roelfsema & van Ooyen [146], Sit & Miikkulainen [147], Golarai *et al.* [207], Reddy *et al.* [208].

The above models made advances in autonomously developing internal states without requiring the human programmer to handcraft the meanings of such states. The primary purpose of the above temporal networks is to predict temporal sequences so that the system can respond similarly when similar sequences are presented again. A major limitation of the above temporal networks is that they do not perform emergent, many-to-one, on-the-fly mapping for all equivalent temporal states. Consequently, they cannot effectively perform brain-like transfer learning discussed below.

J. Networks for Temporal Transfers

The brain does not just do rote learning. It uses logic-like (but numeric) operations in time to transfer temporal associations to other very different temporal sequences, without having to learn every possible temporal sequence exhaustively. A temporal

mechanism that the brain appears to use (Weng [114]) is to autonomously form many equivalent states in early time, similar to the states in a Finite Automaton (FA), but nonsymbolic. Each state considers that multiple temporal experiences are equivalent. Then, the brain automatically applies such equivalences to generate more complex temporal behaviors later, including for many temporal sequences that have never observed. Several well known associative learning procedures belong to this type of temporal transfer learning, such as classical conditioning, secondary classical conditioning, instrumental conditioning, and complex sequence learning—scaffolding from shorter to longer sequences [209].

Almassy *et al.* 1998 [174] designed and experimented with a network for their Darwin robot that demonstrated behaviors similar to the secondary classical conditioning. Zhang & Weng [193] designed and experimented with a network for their SAIL robot that demonstrated a general-purpose nature of associative learning—transfer—transfer shorter and simpler associative skills to new settings and new sequences. This is equivalent to autonomously developing more complex and longer associative mental skills without explicit learning.

The above two networks displayed new skills for tasks that were unknown during the programming time, going beyond a finite combination of otherwise handcrafted symbolic states. However, the full potential of such transfer learning was still unclear till emergent representations could perform general-purpose goal-dependent reasoning as discussed below.

K. Networks for Goal Emergence and Goal-Dependent Reasoning

Connectionist Annette Karmiloff-Smith [210] argued against Fodor's anticonstructivist nativism (similar to symbolic AI) and Piaget's antinativist constructivism (development but without a connectionist account), based on her position that development involves two complementary processes of progressive modularization and progressive explicitation. Her arguments for the two complementary processes is consistent with the principle of emergent representations where the modules are emergent, not having clear cut boundaries, and are experience-dependent. Rogers & McClelland [211] refuted the "theory-theory" approach to semantic knowledge using their feedforward connectionist model. Their language model uses distributed representation (not yet emergent) and allows single-frame classification and text pattern recognition, but their feedforward-only model, without further extension, does not seem to allow recurrent, context-dependent, goal-dependent reasoning.

Newell [3], Pylyshyn [49], Fodor [50] and Harnad [52] proposed that a symbolic system is "rulefully combining". Marvin Minsky [212] argued that prior neural networks do not perform goal-directed reasoning well. These perspectives manifested a great challenge that emergent models have faced for several decades.

Goal-directed in AI typically means "based on the given goal." In this review, I use the term "goal-dependent" since the goal should emerge and change autonomously on the fly.

The SASE model by Weng [213] proposed that internal attention is necessary as internal action (self-effecting) and such internal "thinking" actions are self-aware when they emerge from the motor area. The SASE model was further advanced by the brain-scale but simplified ED network model proposed by Weng [114]. The ED network model is capable of learning concepts and later using concepts as emergent goals for goal-directed reasoning. This was experimentally demonstrated by the Where–What Networks [214], [215], [201], [216], where the goals dynamically emerge and change autonomously, from one time frame to next, among free-viewing, type-goal, and location goal, displaying many different goal patterns for the learned location and type concepts.

The ED network model uses the motor areas in Z as the hubs for emergent concepts (e.g., goal, location and type), abstraction (many forms mapped to one equivalent state), and reasoning (as goal-dependent emergent action). The motor areas appear to be conceptual hubs in the brain, due to:

1) episodic concepts (e.g., object type) can be motorized—verbalized, written, or signed—and be calibrated by a human society;
2) procedure concepts (e.g., arm manipulation) can also be motorized—reaching, grasping, or acting—and be calibrated by a human society.

By calibration above, we mean that feedbacks from human teachers or physical outcome of manipulation can indicate whether the action is sufficiently precise.

Theoretically, the ED network model uses a brain anatomy inspired, general-purpose basic network unit which contains three areas—sensory areas X (e.g., image), internal area Y (e.g., brain), and motor area Z (e.g., muscles). In such a network unit, the Y area connects with other two areas X and Z through two-way connections. The Y area is like a limited-resource "bridge" which predicts the two areas X and Z as its two "banks"—from X and Z at time $t-1$ to X and Z at time $t, t = 1, 2, \ldots$. LCA is a dually optimal model for neurons in the Y area to represent what each best represents. Goal-dependent reasoning-and-thinking behaviors emerge [216], [114] from these highly concise brain-like mechanisms. However, the general-purpose nature of such emergent reasoning mechanisms is unclear without answering the following fundamental question.

L. Can Emergent Models Do All Symbolic Models Can?

As a real-time agent, a hand-crafted FA uses the context dependent rules to reason. However, an FA is static as we discussed above. Can an emergent network incrementally learn to become equivalent to any FA in its external environment based on observing the FA's inputs and outputs?

Theoretically, Weng 2010 [114] explained that, given any FA, an ED can incrementally learn all functions of FA, but using distributed representations (e.g., images) for sensory area X, internal area Y, and motor area Z. The sensory X senses image codes of the symbolic input symbols of the FA. The motor Z inputs (as supervision when needed) and outputs image codes of the states of the FA. The ED generates internal area Y with

its connections (which FA does not have), corresponding to the internal representations.

However, the speed for ED to learn FA was not clear. This situation changed when a generative version of ED, called generative developmental network (GDN) was synthesized along with three theorems [217]. The proofs of the three theorems are available at [218].

Theorem 1 states that there is a GDN that learns any complex imaginary FA, in a grounded way (vector observations), by observing one state transition at a time, immediately, and error free. The FA is represented by the union of different human teachers the GDN met in the life time, but it does not have to be available as a batch at any single time. The imaginary FA does not need any symbolic handcrafting since GDN is grounded. Thanks to grounding, the imaginary FA seems always consistent as long as the grounded sensor of GDN senses sufficient contexts (e.g., different teachers may state different views and the same teacher may state different views on different dates). This is a departure from conventional networks which require iterative approximation, and are subject to local maxima. This means that from any training sequence data that are spatio–temporal in nature, the GDN guarantees error-free in its outputs during the resubstitution tests (i.e., tests using training data) of its incremental learning on the fly.

Theorem 2 states that if the FA-learned GDN computes responses for infinitely many possible sensory inputs and actions in the real physical world but freezes its adaptive part, it generalizes optimally in the sense of maximum likelihood based on its prior FA learning. This means that the above learned GDN is not only error-free for training data, but also optimal for disjoint tests (i.e., tests using data other than the training set).

Theorem 3 states that if the FA-learned GDN is allowed to adapt for infinitely many possible sensory inputs and actions in the real physical world, it "thinks" optimally in the sense of maximum likelihood based on its FA learning and its later learning. This means that the above learned GDN is not only error-free for training data, but also optimal for thinking-like creativity for disjoint tests.

A known limitation of such a motivation-free GDN is that it does not have bias for experience (e.g., likes and dislikes). The motivated versions of the emergent GDN will be reported for different tasks in Daly *et al.* [219] and Paslaski *et al.* [220].

This theory unifies the symbolic models and the emergent models, if one considers the following way: Symbolic models do not have any internal representation and their inputs and outputs are symbolic. Emergent models have internal representations (e.g., the emergent Y area and connections) and their inputs and outputs are grounded in the physical world (e.g., images and muscles). Yes, categorically, it seems that emergent models can do all symbolic models can, at least in principle, but also grounded (i.e., directly sensing and acting in the physical world) and not task-specific.

M. Comments

The advantages of emergent models include:

1) **nonalgorithmic in task space**: neural networks can solve problems that do not yet have an algorithmic solution in the task space or too complex to explicitly handcraft;
2) **numerical in signal space**: neural networks treat all the tasks as regression in signal space, allowing mathematical optimization theories to be applied without even knowing the actual task;
3) **emergence**: representations are fully emergent through fully autonomous DP inside the skull of the network;
4) **uniform processors**: neuronal learning and computations may be carried out in parallel, and relatively low cost hardware with uniform processors can be fabricated which take advantage of this characteristics;
5) **fault tolerance**: partial destruction of a network leads to the corresponding degradation of performance, but other network capabilities are largely retained;
6) **can abstract**: the latest theory of brain-mind network has shown that a network can attend and abstract spatio–temporally;
7) **low, middle, and high levels**: it has been shown that ED networks not only deal with visual attention, perception, cognition, and behavior, but also process logic-like high-level knowledge (e.g., Miyan & Weng [216]), since knowledge hierarchies and their relationships are all emergent;
8) **creativity**: creativity for new knowledge and skills, such as those in Miyan & Weng [216], is from emergent properties of the emergent representations.

At the current stage of knowledge, the existing emergent models still have some shortcomings:

1) **not intuitive**: the responses of neurons do not have linguistically pure meanings, not as intuitive as a symbolic representation;
2) **yet to show scaling up to animal brains**: it has not yet been demonstrated that a large scale emergent model can autonomously develop a wide variety of brain-like mental skills as a cat or a human child, including vision, audition, and language understanding.

However, these two shortcomings are expected to be removed before too long.

V. Conceptual Comparison

In the following, we put the two large schools—the symbolic school and the emergent school—into the same picture for our discussion. Table II gives a summary of some models and their properties. "Mix" denotes symbolic models wherein some local representations are emergent.

A. Brain Architecture

Symbolic AI methods have been motivated by the mind but not tangibly by the brain. They are inspired by psychological observations from human external behaviors, not truly supported by the internal operations of the brain.

The biological work of Sur's group [226] has demonstrated that if the brain was rewired early in life so that the auditory pathway receives visual information, the *visual* representations emerge in the "sound" zone and furthermore, the "sound" zone does some visual tasks successfully. Intuitively, this interesting and illuminating phenomenon seems unlikely to be restricted only to the auditory pathway, since different cortical areas have demonstrated other similar plasticity properties [28], ([34], pp. 270–283). The neuroscience studies showed that in the central

TABLE II
OVERVIEW OF SOME MAJOR AGENT MODELS AND REPRESENTATIONS

Representations and Architectures	Symbolic		Mix	Emergent	World of	Attention		Temporal context	
	Monol.	Contextual			perception	Ascending	Descending	Handcraft	Emergent
3-D map, optimal flow, SLAM	x				real				
FA, HFA, PSG, ACT-R, Soar, CYC, CARUS, Q-learning Net		x			symbolic			x	
Bayesian Net, HMM, MDP, POMDP, Kismet, Cog, HTM		x			real			x	
Koch & Ullman [221], Itti & Koch [222]		x			real	x			
Neuro-Soar, CLARION, LSTM		x	x		real			x	
Schill et al. [223], GNOSIS, ARTSCAN, Albus [42]		x	x		real	x	x	x	
Neocognitron, Cresceptron, SOM, LVQ, LISSOM, ICA, LCA LDA, SVM, HDR, Poggio et al. [175]				x	real				
Elman Net, Jordan Net, Hopfield Net, Boltzman machines, WSA				x	real				x
ARTMAP, Tsotsos et al. [224], AGREL, Deco & Rolls [225], Sit & Miikkulainen [147], SASE				x	real	x	x		
MILN, WWN, ED				x	real	x	x		x

nervous system—from the spinal cord, to the hind brain, the mid brain and the forebrain—neuron in a higher brain area innervate lower brain areas, while a lower area tends to develop earlier [141], [11], [28].

Inspired by the above and other more detailed evidence from neuroscience, the brain-scale model of Weng [114] takes into account five conceptual chunks: **development**—how the brain incrementally grows from experience; **architecture**—how the brain connects and works; **area**—how each brain area emerges and works; **space**—how the brain deals with spatial information; and **time**—how the brain deals with temporal information. From the way these five chunks work intimately together, I argued that none of the five chunks can be missing for understanding how the natural brain-mind works and for simulating the brain-mind using computers and robots. An additional chunk is **modulation**, which deals with actions that depend on dislikes, likes, novelty, confidence, etc. But neuromodulation must be cased on the first five chunks. Thus, the expended brain-mind model is called $5 + 1$ chunk [227].

This model predicts that the internal structure of the internal brain Y emerges from the statistical properties between the two "banks"—sensors and effectors—it serves. Consequently, it is not like a chain of areas proposed by Lee & Mumford [44] for cortical areas. A network of rich cortical connections was reviewed by Stone *et al.* [228], Felleman & Van Essen [9], and Corbetta & Shulman [229]. Weng's model [114] is not completely consistent with the three-role model of Doya [230] who suggested that the basal ganglia, the cerebellum, and the cerebral cortex used supervised, reinforcement, and unsupervised learning modes, respectively. Weng's model predicts that each area Y, regardless in which part of the brain, is capable of performing all kinds of learning depending on whether the teaching signals are generated from its banks (supervised) or the cortical area Y itself (self-learning), and whether the neuromodulaters (e.g., serotonin, dopamine, acetylcholine, and norepinephrine) are transmitted into it or synthesized from it [231], [232] (reinforcement learning, habituation, sensitization, motivation, and beyond). In particular, neuroanatomic wiring [9], [233] does not seem to support that motor error signals (which often require the ground truth) are available in the basal ganglia. However, future work is required to clarify such differences.

B. Symbolic Representations Are External

To relate symbolic models with the brain, we can consider in the following way: When a human designer models a symbolic system using his intuition without looking into the skull of the real brain, he inevitably uses what he observed from the *external* behaviors generated by his brain and other brains. This appears to be what many psychologists are doing. Therefore, the representations they have designed tend to be *external* representations for the brain—externally observed brain behaviors.

For example, Weng [114] explained that a symbolic system is like an FA as its base system. If it uses learning, the learning determines its parameters as probability values, but its "skeleton" base is still *external* representations, since they model the probabilities of human observed inconsistent transitions between every pair of external states.

C. Attention

Spatio–temporal attention seems the essence of natural intelligence, but such skills are accumulated through much experience in the real physical world. For example, perception, cognition, behavior, reasoning, and planning appear to require attention as a central capability: From the "sea" of information externally and internally, context-dependent attention picks up only the information that is directly related to the next action [114]. Perception, cognition, behavior, reasoning, and planning are human words for emphasizing different aspects of this unified real time brain process.

It has been known for many years that descending connections from later areas (e.g., V1) to early cortical areas (e.g., LGN) extensively exist. They do not back-project errors but the back-project signals themselves (Kennedy & Bullier [234],

Perkel *et al.* [235], Felleman & Van Essen [9], Katz & Callaway [236], Salin & Bullier [237], Johnson & Burkhalter [238], Buschman & Miller [239]).

Goal directed search is a major subject in symbolic AI, where the task goal is given from the given task or from a task-specific query. It is static once given. Attention has been largely ignored by the symbolic AI, from computer vision, to speech recognition, and to natural language processing—the three major challenging subjects of AI. Ascending feature fitting has been a major goal of such symbolic recognition systems.

Bottom-up attention has been extensively modeled as saliency during free-viewing [222], [240]–[242]. Free-viewing has been modeled as a special case of more general goal-dependent top-down attention [201] when the top-down attention is "flat."

There have been some neural network models that use descending connections (Grossberg *et al.* [243], [244], Deco & Rolls [225], Roelfsema & van Ooyen [146], Sit & Miikkulainen [147], Weng *et al.* [149]). However, a clear general analysis lacked. In the where–what networks (WWNs) [214], [215], [201], [245] descending information was modeled as goals (e.g., location values and/or type values) that bias not only the outcome of bottom–up competition, but also greatly affect the feature development. The spatio–temporal WWN [201], [200], [246], [114] further models descending information as spatio–temporal context which can represent real-time emerging goal, intent, cognition, and action. Learned WWNs detect and recognize individual objects from natural complex backgrounds [201] and, in language acquisition, generalize beyond learned exemplar sequences (simple thinking) [216]. Any such spatio–temporal context at each time instant is represented as a firing pattern in the motor area. Infinitely many sequences of experience is mapped by the network to a single equivalent spatio–temporal context like a symbolic FA. Such a many-to-one mapping is incrementally learned, representing abstraction and reasoning.

The biological Hebbian mechanism [247], [248] is powerful in that it enables each neuron to incrementally compute the probability of presynaptic firings conditioned on the postsynaptic firing [217]. Using this mechanism, every neuron incrementally figures out the feature vector that it is supposed to detect while it interacts with other connected neurons. However, this learning mechanism seems not sufficient for each neuron to autonomously *disregard* input components that are distractors in the default input field of the neuron (e.g., background pixels outside the contour of a foreground object appearance). A model about biological **synaptic maintenance** by Wang *et al.* [249] shows how each neuron autonomously grows or retracts its synapses. While every neuron does its own synaptic maintenance and Hebbian learning, the entire network emerges to perform general-purpose and highly integrated series of brain-like functions. These functions [114] include, but not limited to, attention, detection, recognition, segmentation, action generation, goal emergence, and reasoning, all directly from unknown complex natural backgrounds, without a need for any handcrafted object models or event models. Handcrafting such models are common in the current computer vision community, using symbolic models (see, e.g., Yuille *et al.* [100] and Yao and Fei-Fei [87]).

In general, emergent models question the widely held view in AI that vision is a self-functioning module which could be isolated from cognition, behavior generation, and motivation. Without actions in the motor areas, the brain cannot figure out where to sense, what is sensed, and what response to generate next. This line of thought seems consistent with many neuroscience and psychological studies (see. e.g., [250]–[252]). A classical experiment of Held & Hein 1963 [253] demonstrated a sharp difference in visual capability between a cat having developed through only passive looking and another one that has developed through a process of autonomous actions.

D. Solution for Inconsistent Behaviors

At any time, a symbolic system may have to choose among multiple inconsistent symbolic actions. The resolution of such actions has been an open problem. The subsumption architecture proposed by R. Brooks [127] and others [254] is such that higher level actions subsume lower levels, where action levels are handcrafted. In many symbolic agents, only one of the actions applicable at any time can be issued based on the highest estimated state-action value [107], [112]. Thus, such systems are not able to learn behaviors that require multiple actions to be carried out concurrently (e.g., after learning single-hand actions, learn two-hand actions).

The ED network model [114] indicates that a resolution to behavior conflicts is a natural consequence of learned experience under the joint context of the sensory ends, the internal area, and the motor ends. For example, parallel location action and type action are produced concurrently in a WWN, but they must be consistent with the current spatio–temporal context.

E. Fully Autonomous Knowledge Self-Update

Using a symbolic expert system, the meanings of each state are only in the mind of the human designer and the machine does not "know" such meanings. Therefore, adding a new piece of knowledge into such a system needs to manually check for consistency with the existing large number of transition rules. Such a consistency checking is tedious and error prone [255].

In contrast, knowledge in an emergent model is represented in a way that facilitates updates. An emergent model not only learns new knowledge (e.g., a new state and the associated transition) but also integrates each piece of new knowledge fully autonomously inside the skull. This is because each internal area ("bridge") connects with all the related neurons in the corresponding areas ("banks"). The well-known bottleneck problem in knowledge updates for a symbolic expert system becomes fully automatic and robust for an emergent network like the ED network.

F. Complexity Difference

Suppose that an agent, natural and artificial, needs to deal with c concepts and each concept takes one of v values. The corresponding symbolic model potentially requires v^c different states, exponential in c. Letting $c = 22$ and $v = 4$, the symbolic model potentially needs $v^c = 4^{22} = (4^2)^{11} = 16^{11} > 10^{11} =$

100 000 000 000, larger than the number of neurons in the human brain. Here are 23 examples of concept: object type, horizontal direction, vertical direction, object pose, apparent scale on the retina, viewing distance, viewing angle, surface texture, surface color, surface reflectance, lighting direction, lighting color, lighting uniformity, material, weight, temperature, deformability, purpose, usage, owner, price, horizontal relationship between two attended objects, vertical relationship between two attended objects. This is a complexity reason why the brian cannot use symbolic states for its internal representations.

We need also consider that a human designer typically merges many states. For example, ACT-R and Soar used handcrafted conditions meant to merge equivalent states into a single meta state. However, it is intractable for a human to examine that many symbolic states and decide which ones are equivalent, especially for an open real world. Therefore, he designs conditions for every meta state without exhaustively checking its validity for the exponential number of real-world states and even more their state trajectories. This is a new complexity reason why symbolic agents are known to be brittle in practical applications. Probabilities only alleviate the inconsistencies of such handcrafted states (better if the states are observable but it is not so with hidden Markov models), but training data are often not sufficient for reaching an acceptable error rates, especially for an open real world.

Next, consider the WWN. For c concepts, each having v values, the number of motor neurons in WWN is only vc. With $v = 4$ and $c = 22$, $vc = 88$ only, instead of 16^{11}. An emergent model like WWN uses further internal weight vectors as clusters in the parallel input space $X \times Z$ from the sensory vector space X and the action vector space Z. The smooth representations in terms of X and Z allow a limited number of synaptic weight clusters to interpolate in the high-dimensional continuous space of $X \times Z$ to automatically partition an unbounded number of states arising from the open-ended real world.

G. Temporal Generalization as Brain-Like Thinking

An FA can be handcrafted to properly deal with temporal context of any finite length, if each symbolic input is fed through time. This is the major reason why an HMM based on FA can recognize spoken sentences of any practical length. But an HMM is handcrafted and symbolic.

An emergent ED-like network can simulate any FA and, furthermore, it is developed incrementally [114]. That is, new representation (i.e., knowledge and skills) are maintained inside the network fully autonomously. In other words, it is now possible to incrementally and autonomously develop a very complex probabilistic FA (e.g., HMM-like or MOPDP-like, but emergent) through sensory and motor interactions.

Furthermore, an FA cannot generalize once handcrafted. HMM and POMDP cannot generalize beyond handling the probability modeled uncertainty either, due to their symbolic nature—there is no distance metrics between the symbolic states.

An emergent ED-like network can perform temporal generalization (brain-like thinking) using internal attention, as shown in member-to-class generalization, member-to-member generalization, subclass-to-superclass generalization in early natural language learning as demonstrated in Miyan & Weng [216]. All such generalization is rooted in temporal thinking-ahead mechanisms of the internal representations, not based on mathematic logic. In particular, knowledge hierarchies, likely more complete than those handcrafted into a hierarchical FA, as well as relationships among knowledge hierarchies are all emergent in an ED.

The ED network model predicts that *brain-like thinking* is not based on handcrafted mathematic logic, but instead based on recursive competitions among neurons in each brain area at every time frame, where every neuron competes to fire using its goodness of match between its weight vector and the two parallel inputs to the area: the descending state-context and the ascending sensory-context.

H. Motivation and Neuromodulation

Motivation seems a more general subject than emotion, by including more basic value-dependent drives, such as dislikes and likes (e.g., pain avoidance and pleasure seeking), and higher value-dependent behaviors, such as novelty seeking. Neuromodulation seems a more general subject than motivation, e.g., by including actions depending on confidence. There have been several models for the intrinsic motivation of a developmental system, to include motivation to learn (Oudeyer *et al.* [256]) and novelty (Huang & Weng [257]) in addition to punishment and rewards dealt with by symbolic reinforcement learning. Almassy *et al.* [174], Krichmar [258] and Krichmar [259] modeled the role of neuromodulators in affecting robot behaviors. Daly *et al.* [219] proposed a framework for neuromodulation based on emergent representations. Such emergent neuromodulation was tested on autonomous navigation settings in Daly *et al.* [219] and on visual recognition in Paslaski *et al.* [220]. Much work remains to integrate the mechanisms of neuromodulation into the capabilities of perception, cognition and behavior of brain-like emergent networks, so that learning is not only fully autonomous inside the skull-closed network but also highly efficient for each living age.

VI. Conclusion

Top–down attention, essential for natural and artificial intelligence, has been largely ignored. Understanding the functions of the natural "genome" program—developmental program (DP)—provides a platform for understanding human intelligence and for solving bottleneck problems of artificial intelligence. AMD requires that the organization of internal representations is fully autonomous, regulated by the DP, natural or artificial. This requirement is not meant to make human programming harder, but to enable autonomous agents to provide brain mechanisms for which the current symbolic models have met intractable limitations—autonomous intent/goal emergence, intent-dependent spatial attention, intent-dependent temporal attention, resolution of inconsistent behaviors, and handcrafting symbolic representations for large practical problems but are not brittle for open task environments. However, the potential of

emergent models needs to be further studied and demonstrated for larger practical problems.

Interestingly, the newly established relationship between each FA and the corresponding ED network indicates that symbolic models and emergent models are intimately related: The FA framework has been used by computer scientists as a tool to communicate about symbolic knowledge since it describes brain's external behaviors using human understandable symbols. An FA is disembodied as it uses symbolic inputs and symbolic outputs. However, a handcrafted FA can be used as a guide in teaching a grounded emergent ED-like network. Therefore brain-like emergent representation includes symbolic representation as a special degenerated case, in the sense that an FA corresponds to some static external behaviors (coded by symbols) of an emergent ED-like network, not including the internal representations and the generalization power of the ED-like network such as thinking for new subjects. The additional power of emergent representations needs to be further explored, including its power and limitation in enabling machines to think like the brain.

REFERENCES

[1] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, pp. 433–460, Oct. 1950.

[2] A. Newell, "Production systems: Models of control structures," in *Visual Information Processing*, W. G. Chase, Ed. New York: Academic, 1973, pp. 283–308.

[3] A. Newell, "Physical symbol systems," *Cogn. Sci.*, vol. 4, no. 4, pp. 135–183, 1980.

[4] A. Allport, "Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years," in *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*, D. E. Meyer and S. Kornblum, Eds. Cambridge, MA: MIT Press, 1993, pp. 111–132.

[5] P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cogn. Syst. Res.*, vol. 10, pp. 141–160, 2009.

[6] A. Oreback and H. I. Christensen, "Evaluation of architectures for mobile robots," *Autonom. Robot.*, vol. 14, pp. 33–49, 2003.

[7] D. Vernon, G. Metta, and G. Sandini, "A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents," *IEEE Trans. Evol. Comput.*, vol. 11, no. 2, pp. 151–180, Jun. 2007.

[8] L. W. Barsalou, "Grounded cognition," *Annu. Rev. Psychol.*, vol. 59, pp. 617–645, 2008.

[9] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cereb. Cortex*, vol. 1, pp. 1–47, 1991.

[10] D. C. Van Essen, J. W. Lewis, H. A. Drury, N. Hadjikhani, R. B. H. Tootell, M. Bakircioglu, and M. I. Miller, "Mapping visual cortex in monkeys and human using surface-based atlases," *Vis. Res.*, vol. 41, pp. 1359–1378, 2001.

[11] *Principles of Neural Science*, E. R. Kandel, J. H. Schwartz, and T. M. Jessell, Eds., 4th ed. New York: McGraw-Hill, 2000.

[12] J. Weng and J. McClelland, "Dialog initiation: How the mind works and how the brain develops," *IEEE CIS Autonom. Mental Develop. Newsletter*, vol. 4, no. 2, p. 5, 2007.

[13] R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford, U.K.: Oxford Univ. Press, 1989.

[14] R. Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford, U.K.: Oxford Univ. Press, 1994.

[15] *Artificial General Intelligence*, B. Goertzel and C. Pennachin, Eds. Berlin, Germany: Springer-Verlag, 2007.

[16] D. George and J. Hawkins, "Towards a mathematical theory of cortical micro-circuits," *PLoS Comput. Biolo.*, vol. 5, no. 10, pp. 1–26, 2009.

[17] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, 1980.

[18] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, no. 6, pp. 1233–1258, 1987.

[19] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[20] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, "Autonomous mental development by robots and animals," *Science*, vol. 291, no. 5504, pp. 599–600, 2001.

[21] J. Weng and I. Stockman, "Autonomous mental development: Workshop on development and learning," *AI Mag.*, vol. 23, no. 2, pp. 95–98, 2002.

[22] S. Oyama, *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd ed. Durham, NC: Duke Univ. Press, 2000.

[23] W. M. Rideout, K. Eggan, and R. Jaenisch, "Nuclear cloning and epigenetic reprogramming of the genome," *Science*, vol. 293, no. 5532, pp. 1093–1098, 2001.

[24] W. Reik, W. Dean, and J. Walter, "Epigenetic reprogramming in mammalian development," *Science*, vol. 293, no. 5532, pp. 1089–1093, 2001.

[25] M. A. Surani, "Reprogramming of genome function through epigenetic inheritance," *Nature*, vol. 414, pp. 122–128, 2001.

[26] E. Li, "Chromatin modification and epigenetic reprogramming in mammanlian development," *Nature Rev. Genet.*, vol. 3, no. 9, pp. 662–673, 2002.

[27] L. C. Karz and C. J. Shatz, "Synaptic activity and the construction of cortical circuits," *Science*, vol. 274, no. 5290, pp. 1133–1138, 1996.

[28] M. Sur and J. L. R. Rubenstein, "Patterning and plasticity of the cerebral cortex," *Science*, vol. 310, pp. 805–810, 2005.

[29] M. Cole and S. R. Cole, *The Development of Children*, 3rd ed. New York: Freeman, 1996.

[30] J. Weng, N. Ahuja, and T. S. Huang, "Cresceptron: A self-organizing neural network which grows adaptively," in *Proc. Int. Joint Conf. Neural Netw.*, Baltimore, MD, Jun. 1992, vol. 1, pp. 576–581.

[31] J. Weng, N. Ahuja, and T. S. Huang, "Learning recognition and segmentation of 3-D objects from 2-D images," in *Proc. IEEE 4th Int. Conf. Comput. Vis.*, May 1993, pp. 121–128.

[32] J. Weng, N. Ahuja, and T. S. Huang, "Learning recognition and segmentation using the cresceptron," *Int. J. Comput. Vis.*, vol. 25, no. 2, pp. 109–143, Nov. 1997.

[33] J. L. McClelland, "The interaction of nature and nurture in development: A parallel distributed processing perspective," in *International Perspectives on Psychological Science*, P. Bertelson, P. Eelen, and G. d'Ydewalle, Eds. Hillsdale, NJ: Erlbaum, 1994, vol. 1, pp. 57–88.

[34] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett, *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press, 1997.

[35] S. Quartz and T. J. Sejnowski, "The neural basis of cognitive development: A constructivist manifesto," *Behav. Brain Sci.*, vol. 20, no. 4, pp. 537–596, 1997.

[36] J. Elman, "Learning and development in neural networks: The importance of starting small," *Cognition*, vol. 48, no. 1, pp. 71–99, 1993.

[37] J. Pearl, "Fusion, propagation, and structuring in belief nentworks," *Artif. Intell.*, vol. 29, pp. 241–288, 1986.

[38] M. I. Jordan and C. Bishop, "Neural networks," in *CRC Handbook of Computer Science*, A. B. Tucker, Ed. Boca Raton, FL: CRC Press, 1997, pp. 536–556.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] R. Sun, E. Merrill, and T. Peterson, "From implicit skills to explicit knowledge: A bottom-up model of skill learning," *Cogn. Sci.*, vol. 25, pp. 203–244, 2001.

[41] R. Sun, "The importance of cognitive architectures: An analysis based on CLARION," *J. Exp. Theoret. Artif. Intell.*, vol. 19, pp. 159–193, 2007.

[42] J. S. Albus, "A model of computation and representation in the brain," *Inform. Sci.*, vol. 180, no. 9, pp. 1519–1554, 2010.

[43] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, "How to grow a mind: Statistics, structure, and abstraction," *Science*, vol. 331, pp. 1279–1285, 2011.

[44] T. S. Lee and D. Mumford, "Hierarchical Bayesian inference in the visual cortex," *J. Opt. Soc. Amer. A*, vol. 20, no. 7, pp. 1434–1448, 2003.

[45] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: A survey," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, pp. 12–34, 2009.

[46] T. Froese and T. Ziemke, "Enactive artificial intelligence: Investigating the systemic organization of life and mind," *Artif. Intell.*, vol. 173, pp. 466–500, 1009.

[47] J. Weng and W. Hwang, "From neural networks to the brain: Autonomous mental development," *IEEE Comput. Intell. Mag.*, vol. 1, no. 3, pp. 15–31, 2006.

[48] J. Weng, "Task muddiness, intelligence metrics, and the necessity of autonomous mental development," *Minds Mach.*, vol. 19, no. 1, pp. 93–115, 2009.

[49] Z. W. Pylyshyn, "Computation and cognition: Issues in the foundations of cognitive science," *Behav. Brain Sci.*, vol. 3, pp. 111–132, 1980.

[50] J. A. Fodor, "Précis of the modularity of mind," *Behav. Brain Sci.*, vol. 8, pp. 1–5, 1985.

[51] S. Harnad, *Categorical Perception: The Groundwork of Cognitition*. New York: Cambridge Univ. Press, 1987.

[52] S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335–346, 1990.

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

[54] U. R. Dhond and J. K. Aggarwal, "Structure from stereo: A review," *IEEE Trans. Systems, Man, Cybernet.*, vol. 19, no. 6, pp. 1489–1510, Nov.–Dec. 1989.

[55] J. Weng, P. Cohen, and N. Rebibo, "Motion and structure estimation from stereo image sequences," *IEEE Trans. Robot. Autom.*, vol. 8, no. 3, pp. 362–382, Jun. 1992.

[56] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 675–684, Jul. 2000.

[57] W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*. Cambridge, MA: MIT Press, 1981.

[58] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. New York: Freeman, 1982.

[59] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, Aug. 1981.

[60] B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On benchmarking optical flow," *Comput. Vis. Image Understand.*, vol. 84, no. 1, pp. 765–773, 2001.

[61] J. Weng, "Image matching using the windowed Fourier phase," *Int. J. Comput. Vis.*, vol. 11, no. 3, pp. 211–236, 1993.

[62] H. C. Longuet-Higgins, "A computer program for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, 1981.

[63] R. Cipolla, Y. Okamoto, and Y. Kuno, "Robust structure from motion using motion parallax," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.ion*, 1993, pp. 374–382.

[64] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of 3-d motion parameters of rigid bodies with curved surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 13–27, Jun. 1984.

[65] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two perspective views: Algorithm, error analysis and error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 5, pp. 451–476, May 1989.

[66] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, 1992.

[67] J. O'Rourke and N. Badler, "Model-based image analysis of human motion using constraint propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 6, pp. 523–536, Jun. 1980.

[68] A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 730–742, 1991.

[69] P. Cheeseman and P. Smith, "On the representation and estimation of spatial uncertainty," *Int. J. Robot.*, vol. 5, pp. 56–68, 1986.

[70] J. Weng, Y. Cui, and N. Ahuja, "Transitory image sequences, asymptotic properties, and estimation of motion and structure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 451–464, May 1997.

[71] M. Dissanayaka, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localisation and map building problem," *IEEE Trans. Robot. Autom.*, vol. 17, p. 229241, 2001.

[72] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (SLAM): Part I the essential algorithms,," *IEEE Trans. Robot.*, vol. 17, p. 99110, 2006.

[73] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (SLAM): Part II state of the art,," *IEEE Trans. Robot.*, vol. 17, p. 99110, 2006.

[74] D. Hähnel, D. Fox, W. Burgard, and S. Thrun, "A highly efficient FastSLAM algorithm for generating cyclic maps of large-scale environments from raw laser range measurements," in *Proc. Conf. Intell. Robot. Syst. (IROS)*, 2003.

[75] M. Montemerlo, B. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proc. AAAI Nat. Conf. Artif. Intell.*, Edmonton, AB, Canada, 2002, AAAI.

[76] Z. Liu, Z. Zhang, C. Jacobs, and M. Cohen, "Rapid modeling of animated faces from vide," in *Proc. 3rd Int. Conf. Vis. Comput.*, Mexico City, Mexico, Sep. 18–22, 2000, pp. 58–67.

[77] W. E. L. Grimson and T. Lozano-Perez, "Model-based recognition and localization from sparse range or tactile data," *Int. J. Robot. Res.*, vol. 3, no. 3, pp. 3–35, 1984.

[78] K. Ikeuchi and T. Kanade, "Automatic generation of object recognition programs," *Proc. IEEE*, vol. 76, no. 8, pp. 1016–1035, Aug. 1988.

[79] A. K. Jain and R. L. Hoffman, "Evidence-based recognition of 3-D objects,," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 6, pp. 783–802, Jun. 1988.

[80] Y. Lamdan and H. J. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," in *Proc. IEEE 2nd Int. Conf. Comput. Vis.*, 1988, pp. 238–246.

[81] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance,," *Int. J. Comput. Vis.*, vol. 14, no. 1, pp. 5–24, Jan. 1995.

[82] Y. Cui and J. Weng, "View-based hand segmentation and hand-sequence recognition with complex backgrounds," in *Proc. Int. Conf. Pattern Recognit.*, Vienna, Austria, Aug. 25–30, 1996, pp. 617–621.

[83] D. Roy and A. Pentland, "Learning words from sights and sounds: A computational model," *Cogn. Sci.*, vol. 26, no. 1, pp. 113–146, 2002.

[84] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1775–1789, Oct. 2009.

[85] A. Bobick and A. Wilson, "A state-based technique for the summarization and recognition of gesture," in *Proc. 5th Int. Conf. Comput. Vis.*, Boston, MA, 1995, pp. 382–388.

[86] T. Darrell and A. Pentland, "Active gesture recognition using partial observable Markov decision processes," in *Proc. 10th Int. Conf. Pattern Recognit.*, Vienna, Austria, 1996.

[87] B. Yao and L. Fei-Fei, "Modeling mutual context of object and human pose in human-object interaction activities," in *Proc. Comput. Vis. Pattern Recognit.*, San Francisco, CA, Jun. 15–17, 2010, pp. 1–8.

[88] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.

[89] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[90] R. Brooks, "Intelligence without reason," in *Proc. Int. Joint Conf. Artif. Intell.*, Sydney, Australia, Aug. 1991, pp. 569–595.

[91] Y. Aloimonos, "Purposive active vision," *Comput. Vis., Graphics, Image Process.: Image Understand.*, vol. 17, no. 1–3, pp. 285–348, Aug. 1992.

[92] J. E. Hopcroft, R. Motwani, and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*. Boston, MA: Addison-Wesley, 2006.

[93] J. R. Anderson, *Rules of the Mind*. Mahwah, NJ: Lawrence Erlbaum, 1993.

[94] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: An architecture for general intelligence," *Artif. Intell.*, vol. 33, pp. 1–64, 1987.

[95] B. Cho, P. S. Rosenbloom, and C. P. Dolan, "Neuro-soar: A neural-network architecture for goal-oriented behavior," in *The Soar Papers*, P. S. Rosenbloom, J. E. Laird, and A. Newell, Eds. Cambridge, MA: MIT Press, 1993, pp. 1199–1203.

[96] J. E. Laird, E. S. Yager, and C. M. Tuck, "Robo-soar: An integration of external interaction, planning, and learning using soar," *Robot. Autonom. Syst.*, vol. 8, pp. 113–129, 1991.

[97] D. B. Lenat, "CYC: A large-scale investment in knowledge infrastructure," *Commun. ACM*, vol. 38, no. 11, pp. 33–38, 1995.

[98] T. Starner and A. Pentland, "Visual recognition of American sign language using hidden Markov models," in *Proc. Int. Workshop Automatic Face Gesture Recognit.*, Zurich, Switzerland, Jun. 1995, pp. 189–194.

[99] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2003.

[100] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 113–140, 2005.

[101] L. J. Li and L. Feifei, "What, where and who? classifying events by scene object recognition," in *IEEE Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, Oct. 14–20, 2007.
[102] M. L. Puterman, *Markov Decision Processes*. New York: Wiley, 1994.
[103] L. R. Rabiner, L. G. Wilpon, and F. K. Soong, "High performance connected digit recognition using hidden Markov models," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 8, pp. 1214–1225, Aug. 1989.
[104] W. S. Lovejoy, "A survey of algorithmic methods for partially observed Markov decision processes," *Ann. Operations Res.*, vol. 28, pp. 47–66, 1991.
[105] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[106] L. R. Rabiner, "Toward vision 2001: Voice and audio processing considerations," *AT&T Techn. J.*, vol. 74, no. 2, pp. 4–13, 1995.
[107] C. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, 1992.
[108] L. Lin, "Self-improving reactive agents based on reinforcement learning, planning and teaching," *Mach. Learn.*, vol. 8, no. 3–4, pp. 293–322, 1992.
[109] A. Schwartz, "A reinforcement learning method for maximizing undiscounted rewards," in *Proc. Int. Joint Conf. Artif. Intell.*, Chambery, France, 1993, pp. 289–305.
[110] S. Mahadevan, "Average reward reinforcement learning: Foundation, algorithms, and empirical results," *Mach. Learn.*, vol. 22, pp. 159–196, 1996.
[111] M. J. Mataric, "Reinforcement learning in the multi-robot domain," *Autonom. Robot.*, vol. 4, no. 1, pp. 73–83, Jan. 1997.
[112] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, 1996.
[113] A. Barto, S. Singh, and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills," in *Proc. 3rd Int. Conf. Develop. Learn.*, 2004, pp. 1–8.
[114] J. Weng, "A 5-chunk developmental brain-mind network model for multiple events in complex backgrounds," in *Proc. Int. Joint Conf. Neural Netw.*, Barcelona, Spain, Jul. 18–23, 2010, pp. 1–8.
[115] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychol. Rev.*, vol. 111, no. 4, p. 103601060, 2004.
[116] A. M. Nuxoll and J. E. Laird, "Extending cognitive architecture with episodic memory," in *Proc. 22nd AAAI Conf. Artif. Intell.*, Vancouver, BC, Canada, Jul. 22–26, 2007.
[117] P. Langley, K. Cummings, and D. Shapiro, "Hierarchical skills and cognitive architectures," in *Proc. 26th Annu.Conf. Cogn. Sci. Soc.*, Chicago, IL, 2004, pp. 779–784.
[118] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[119] J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. New York: Macmillan, 1993.
[120] J. White, *Markov Decision Processes*. Chichester: Wiley, 1993.
[121] A. Waibel and K. Lee, *Readings in Speech Recognition*. San Mateo, CA: Morgan Kaufmann, 1990.
[122] I. R. Alexander, G. H. Alexander, and K. F. Lee, "Survey of current speech technology," *Commun. ACM*, vol. 37, no. 3, pp. 52–57, Mar. 1994.
[123] J. Y. J. Ohya and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1992, pp. 379–385.
[124] T. Darrell and A. Pentland, "Space-time gesture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, Jun. 1993, pp. 335–340.
[125] S. Koenig and R. Simmons, "A robot navigation architecture based on partially observable Markov decision process models," in *Artificial Intelligence Based Mobile Robotics: Case Studies of Successful Robot Systems*, D. Kortenkamp, R. Bonasso, and R. Murphy, Eds. Cambridge, MA: MIT Press, 1998, pp. 91–122.
[126] G. Theocharous and S. Mahadevan, "Approximate planning with hierarchical partially observable Markov decision processes for robot navigation," in *IEEE Conf. Robot. Autom.*, Washington, DC, 2002.
[127] R. A. Brooks, "A robust layered control system for a mobile robot," *IEEE J. Robot. Autom.*, vol. 2, no. 1, pp. 14–23, Mar. 1986.
[128] R. A. Brooks, "Intelligence without representation," *Artif. Intell.*, vol. 47, pp. 139–160, 1991.
[129] C. Breazeal, "Social constraints on animate vision," in *Proc. IEEE Int. Conf. Humanoid Robot.*, Cambridge, MA, Sep. 7–8, 2000.
[130] C. Breazeal and B. Scassellati, "Infant-like social interactions between a robot and a human caretaker," *Adapt. Behav.*, vol. 8, pp. 49–74, 2000.
[131] D. H. Ballard and C. M. Brown, "Principles of animate vision," *Comput. Vis., Graphics, Image Process.: Image Understand.*, vol. 56, no. 1, pp. 3–21, Jul. 1992.
[132] F. Wörgötter, A. Agostini, N. Krger, N. Shylo, and B. Porr, "Cognitive agents a procedural perspective relying on the predictability of object? action? complexes (OACs)," *Robotics and Autonom. Syst.*, vol. 57, no. 4, pp. 420–432, 2009.
[133] A. G. Cohn, D. C. Hogg, B. Bennett, V. Devin, A. Galata, D. R. Magee, C. Needham, and P. Santos, "Cognitive vision: Integrating symbolic qualitative representations with computer vision," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches, LNCS*, H. Christensen and H.-H. Nagel, Eds. Heidelberg, Germany: Springer-Verlag, 2005, pp. 211–234.
[134] D. Vernon, "Cognitive vision: The case for embodied perception," *Image Vis. Comput.*, vol. 26, no. 1, pp. 127–140, 2008.
[135] J. Taylor, M. Hartley, N. Taylor, C. Panchev, and S. Kasderidis, "A hierarchical attention-based neural network architecture, based on human brain guidance, for perception, conceptualisation, action and reasoning," *Image Vis. Comput.*, vol. 27, no. 11, pp. 1641–1657, 2009.
[136] A. Fazl, S. Grossberg, and E. Mingolla, "View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds," *Cogn. Psychol.*, vol. 58, pp. 1–48, 2009.
[137] J. Weng and S. Chen, "Visual learning with navigation as an example," *IEEE Intell. Syst.*, vol. 15, pp. 63–71, Sep./Oct. 2000.
[138] F. H. Hsu, "IBM's deep blue chess grandmaster chips," *IEEE Micro*, vol. 19, no. 2, pp. 70–81, 1999.
[139] M. Sur and C. A. Leamey, "Development and plasticity of cortical areas and networks," *Nature Rev. Neurosci.*, vol. 2, pp. 251–262, 2001.
[140] Y. Munakata and J. L. McClelland, "Connectionist models of development," *Develop. Sci.*, vol. 6, no. 4, pp. 413–429, 2003.
[141] W. K. Purves, D. Sadava, G. H. Orians, and H. C. Heller, *Life: The Science of Biology*, 7th ed. Sunderland, MA: Sinauer, 2004.
[142] S. F. Gilbert, *Developmental Biology*, 8th ed. Sunderland, MA: Sinauer, 2006.
[143] , J. L. McClelland and D. E. Rumelhart, Eds., *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986, vol. 2.
[144] O. Sporns, N. Almassy, and G. Edelman, "Plasticity in value systems and its role in adaptive behavior," *Adapt. Behav.*, vol. 7, no. 3, 1999.
[145] J. C. Wiemer, "The time-organized map algorithm: Extending the self-organizing map to spatiotemporal signals," *Neural Comput.*, vol. 15, pp. 1143–1171, 2003.
[146] P. R. Roelfsema and A. van Ooyen, "Attention-gated reinforcement learning of internal representations for classification," *Neural Computat.*, vol. 17, pp. 2176–2214, 2005.
[147] Y. F. Sit and R. Miikkulainen, "Self-organization of hierarchical visual maps with feedback connections," *Neurocomput.*, vol. 69, pp. 1309–1312, 2006.
[148] G. E. Hinton, "Learning multiple layers of representation," *Trend. Cogn. Sci.*, vol. 11, no. 10, pp. 428–434, 2007.
[149] J. Weng, T. Luwang, H. Lu, and X. Xue, "Multilayer in-place learning networks for modeling functional layers in the laminar cortex," *Neural Netw.*, vol. 21, pp. 150–159, 2008.
[150] C. Blakemore and G. F. Cooper, "Development of the brain depends on the visual environment," *Nature*, vol. 228, pp. 477–478, Oct. 1970.
[151] M. M. Merzenich, J. H. Kaas, J. T. Wall, M. Sur, R. J. Nelson, and D. J. Felleman, "Progression of change following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys," *Neuroscience*, vol. 10, no. 3, pp. 639–665, 1983.
[152] M. M. Merzenich, R. J. Nelson, M. P. Stryker, M. S. Cynader, A. Schoppmann, and J. M. Zook, "Somatosensory cortical map changes following digit amputation in adult monkeys," *J. Compar. Neurol.*, vol. 224, pp. 591–605, 1984.
[153] E. M. Callaway and L. C. Katz, "Effects of binocular deprivation on the development of clustered horizontal connections in cat striate cortex," in *Proc. Nat. Acad. Sci.*, 1991, vol. 88, no. 3, pp. 745–749.
[154] C. D. Gilbert and T. N. Wiesel, "Receptive field dynamics in adult primary visual cortex," *Nature*, vol. 356, pp. 150–152, 1992.
[155] S. Lowel and W. Singer, "Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity," *Science*, vol. 255, pp. 209–212, 1992.

[156] M. B. Feller, D. P. Wellis, D. Stellwagen, F. S. Werblin, and C. J. Shatz, "Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves," *Science*, vol. 272, no. 5265, pp. 1182–1187, 1996.

[157] M. Meister, R. O. L. Wong, D. A. Baylor, and C. J. Shatz, "Synchronous burts of action-potentials in the ganglion cells of the developing mammalian retina," *Science*, vol. 252, pp. 939–943, 1991.

[158] X. Wang, M. M. Merzenich, K. Sameshima, and W. M. Jenkins, "Remodeling of hand representation in adult cortex determined by timing of tactile stimulation," *Nature*, vol. 378, no. 2, pp. 13–14, 1995.

[159] V. Dragoi and M. Sur, "Plasticity of orientation processing in adult visual cortex," in *Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds. Cambridge, MA: MIT Press, 2004, pp. 1654–1664.

[160] T. Hosoya, S. A. Baccus, and M. Meister, "Dynamic predictive coding by the retina," *Nature*, vol. 436, pp. 71–77, 2005.

[161] T. Bonhoeffer and A. Grinvald, "ISO-orientation domains in cat visual cortex are arranged in pinwheel-like patterns," *Nature*, vol. 353, pp. 429–431, 1991.

[162] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol.Cybernet.*, vol. 36, pp. 193–202, 1980.

[163] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Berlin, Germany: Springer-Verlag, 2001.

[164] T. Kohonen, *Self-Organizing Maps*, 2nd ed. Berlin, Germany: Springer-Verlag, 1997.

[165] N. Zhang and J. Weng, "A developing sensory mapping for robots," in *Proc. IEEE 2nd Int. Conf. Develop. Learn. (ICDL 2002)*, Cambridge, MA, Jun. 12–15, 2002, pp. 13–20, MIT.

[166] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh, *Computational Maps in the Visual Cortex*. Berlin, Germany: Springer-Verlag, 2005.

[167] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[168] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vis. Res.*, vol. 37, no. 23, pp. 3327–3338, 1997.

[169] A. Hyvarinen, "Survey on independent component analysis," *Neural Comput. Surveys*, vol. 2, pp. 94–128, 1999.

[170] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[171] C. von der Malsburg, "Self-organization of orientation-sensitive cells in the striate cortex," *Kybernetik*, vol. 15, pp. 85–100, 1973.

[172] D. Field, "What is the goal of sensory coding?," *Neural Comput.*, vol. 6, pp. 559–601, 1994.

[173] K. Obermayer, H. Ritter, and K. J. Schulten, "A neural network model for the formation of topographic maps in the CNS: Development of receptive fields," in *Proc. Int. Joint Conf. Neural Netw.*, San Diego, C, 1990, vol. 2, pp. 423–429.

[174] N. Almassy, G. M. Edelman, and O. Sporns, "Behavioral constraints in the development of neural properties: A cortical model embedded in a real-world device," *Cereb. Cortex*, vol. 8, no. 4, pp. 346–361, 1998.

[175] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, 2007.

[176] A. S. Reber, S. M. Kassin, S. Lewis, and G. Cantor, "On the relationship between implicit and explicit modes in the learning of a complex rule structure," *J. Exp. Psychol.: Human Learn. Memory*, vol. 6, no. 5, pp. 492–502, 1980.

[177] R. Sun, P. Slusarz, and C. Terry, "The interaction of the explicit and the implicit in skill learning: A dual-process approach," *Psychol. Rev.*, vol. 112, no. 1, p. 59192, 2005.

[178] J. Weng and N. Zhang, "Optimal in-place learning and the lobe component analysis," in *Proc. IEEE World CongressComput. Intell.*, Vancouver, BC, Canada, Jul. 16–21, 2006, pp. 1–8.

[179] P. J. Werbos, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. Chichester: Wiley, 1994.

[180] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[181] S. E. Fahlman and C. Lebiere, The Cascade-Correlation Learning Architecture School of Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-90-100, Feb. 1990.

[182] T. R. Shultz, *Computational Developmental Psychology*. Cambridge, MA: MIT Press, 2003.

[183] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

[184] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Atlanta, GA, May 1994, pp. 2148–2151.

[185] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[186] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[187] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*. New York: Wiley, 1998.

[188] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[189] Y. Lin, "Support vector machines and the bayes rule in classification," *Data Mining Knowledge Discov.*, vol. 6, pp. 259–275, 2002.

[190] W. S. Hwang and J. Weng, "Hierarchical discriminant regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1277–1293, Nov. 2000.

[191] J. Weng and W. Hwang, "Incremental hierarchical discriminant regression," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 397–415, Apr. 2007.

[192] Y. Zhang, J. Weng, and W. Hwang, "Auditory learning: A developmental method," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 601–616, Jun. 2005.

[193] Y. Zhang and J. Weng, "Task transfer by a developmental robot," *IEEE Trans. Evol. Comput.*, vol. 11, no. 2, pp. 226–248, Apr. 2007.

[194] Y. Zhang and J. Weng, "Multimodal developmental learning," *IEEE Trans. Autonom. Mental Develop.*, vol. 2, no. 3, pp. 149–166, Sep. 2010.

[195] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural networks," *Neural Netw.*, vol. 4, pp. 565–588, 1991.

[196] J. Weng and M. D. Luciw, "Optimal in-place self-organization for cortical development: Limited cells, sparse coding and cortical topography," in *Proc. 5th Int. Conf. Develop. Learn. (ICDL'06)*, Bloomington, IN, May 31–Jun. 2006, pp. 1–7.

[197] J. Weng, H. Lu, T. Luwang, and X. Xue, "In-place learning for positional and scale invariance," in *Proc. IEEE World Congress Comput. Intell.*, Vancouver, BC, Canada, July 16–21, 2006.

[198] J. Weng, T. Luwang, H. Lu, and X. Xue, "A multilayer in-place learning network for development of general invariances," *Int. J. Human. Robot.*, vol. 4, no. 2, pp. 281–320, 2007.

[199] M. Luciw and J. Weng, "Top-down connections in self-organizing Hebbian networks: Topographic class grouping," *IEEE Trans. Autonom. Mental Develop.*, vol. 2, no. 3, pp. 248–261, Sep. 2010.

[200] J. Weng, Q. Zhang, M. Chi, and X. Xue, "Complex text processing by the temporal context machines," in *Proc. IEEE 8th Int. Conf. Develop. Learn.*, Shanghai, China, Jun. 4–7, 2009, pp. 1–8.

[201] M. Luciw and J. Weng, "Where What Network 3: Developmental top-down attention with multiple meaningful foregrounds," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Barcelona, Spain, Jul. 18–23, 2010, pp. 4233–4240.

[202] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. Nat. Acad. Sci. USA*, 1982, vol. 79, no. 8, pp. 2554–2558.

[203] H. Cruse, *Neural Networks as Cybernetic Systems*, 2nd ed. Bielefeld, Germany: , 2006.

[204] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for boltzmann machines," *Cogn. Sci.*, vol. 9, pp. 147–169, 1985.

[205] G. E. Hinton, S. Osindero, and Y.-. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, pp. 1527–1554, 2006.

[206] C. W. Omlin and C. L. Giles, "Constructing deterministic finite-state automata in recurrent neural networks," *J. ACM*, vol. 43, no. 6, pp. 937–972, 1996.

[207] G. Golarai, D. G. Ghahremani, S. Whitfield-Gabrieli, A. Reiss, J. L. Eberhard, J. D. E. Gabrieli, and K. Grill-Spector, "Differential of high-level visual cortex correlates with category-specific recognition memory," *Nature Neurosci.*, vol. 10, no. 4, pp. 512–522, 2007.

[208] L. Reddy, F. Moradi, and C. Koch, "Top-down biases win against focal attention in the fusiform face area," *Neuroimage*, vol. 38, pp. 730–739, 2007.

[209] M. Domjan, *The Principles of Learning and Behavior*, 4th ed. Belmont, CA: Brooks/Cole, 1998.

[210] A. Karmiloff-Smith, "Précis of of beyond modularity: A developmental perspective on cognitive science," *Behav. Brain Sci.*, vol. 17, pp. 693–707, 1994.
[211] T. T. Rogers and J. L. McClelland, "Preecis of semantic cognition: A parallel distributed processing approach," *Behav. Brain Sci.*, vol. 31, pp. 689–749, 2008.
[212] M. Minsky, "Logical versus analogical or symbolic versus connectionist or neat versus scruffy," *AI Mag.*, vol. 12, no. 2, pp. 34–51, 1991.
[213] J. Weng, "On developmental mental architectures," *Neurocomputing*, vol. 70, no. 13–15, pp. 2303–2323, 2007.
[214] Z. Ji, J. Weng, and D. Prokhorov, "Where-what network 1: "Where" and "What" assist each other through top-down connections," in *Proc. IEEE Int. Conf. Develop. Learn.*, Monterey, CA, Aug. 9–12, 2008, pp. 61–66.
[215] Z. Ji and J. Weng, "WWN-2: A biologically inspired neural network for concurrent visual attention and recognition," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Barcelona, Spain, Jul. 18–23, 2010, pp. 1–8.
[216] K. Miyan and J. Weng, "WWN-Text: Cortex-like language acquisition with What and Where," in *Proc. IEEE 9th Int. Conf. Devel. Learn.*, Ann Arbor, MI, Aug. 18–21, 2010, pp. 280–285.
[217] J. Weng, "Three theorems: Brain-like networks logically reason and optimally generalize," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, Jul. 31–Aug. 5 2011, pp. 1–8.
[218] J. Weng, Three Theorems About Developmental Networks and the Proofs Dept. Comput. Sci., Michigan State Univ., East Lansing, MI, Tech. Rep. MSU-CSE-11-9, May 12, 2011.
[219] J. Daly, J. Brown, and J. Weng, "Neuromorphic motivated systems," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, Jul. 31–August 5 2011, pp. 1–8.
[220] S. Paslaski, C. VanDam, and J. Weng, "Modeling dopamine and serotonin systems in a visual recognition network," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, Jul. 31–Aug. 5 2011, pp. 1–8.
[221] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiol.*, vol. 4, pp. 219–227, 1985.
[222] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[223] K. Schill, E. Umkehrer, S. Beinlich, G. Krieger, and C. Zetzsche, "Scene analysis with saccadic eye movemens: Top-down and bottom-up modeling," *J. Electron. Imaging*, vol. 10, no. 1, pp. 152–160, 2001.
[224] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artif. Intell.*, vol. 78, pp. 507–545, 1995.
[225] G. Deco and E. T. Rolls, "A neurodynamical cortical model of visual attention and invariant object recognition," *Vis. Res.*, vol. 40, pp. 2845–2859, 2004.
[226] J. Sharma, A. Angelucci, and M. Sur, "Induction of visual orientation modules in auditory cortex," *Nature*, vol. 404, pp. 841–847, 2000.
[227] J. Weng, "Through the symbol-grounding problem see the two largest hindrances of science," *AMD Newsletter*, vol. 8, no. 1, pp. 10–13, 2011.
[228] J. Stone, B. O. Dreher, and A. Leventhal, "Hierarchical and parallel mechanisms in the organization of visual cortex," *Brain Res. Rev.*, vol. 1, pp. 345–394, 1979.
[229] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Rev. Neural Sci.*, vol. 3, pp. 201–215, 2002.
[230] K. Doya, "Complementary roles of basal ganglia and cerebellum in learning and motor control," *Current Opinion Neurobiol.*, vol. 10, no. 6, pp. 732–739, 2000.
[231] J. Balthazart and G. F. Ball, "Is brain estradiol a hormone or a neurotransmitter?," *Trend. Neurosci.*, vol. 29, no. 5, pp. 241–249, May 2006.
[232] L. Remage-Healey, M. J. Coleman, R. K. Oyama, and B. A. Schlinger, "Brain estrogens rapidly strengthen auditory encoding and guide song preference in a songbird," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 8, pp. 3852–3857, 2010.
[233] E. M. Callaway, "Feedforward, feedback and inhibitory connections in primate visual cortex," *Neural Netw.*, vol. 17, pp. 625–632, 2004.
[234] H. Kennedy and J. Bullier, "A double-labelling investigation of the afferent connectivity to cortical areas v1 and v2 of the macaque monkey," *J. Neurosci.*, vol. 5, no. 10, pp. 2815–2830, 1985.
[235] D. J. Perkel, J. Bullier, and H. Kennedy, "Topography of the afferent connectivity of area 17 of the macaque monkey," *J. Comp. Neurosci.*, vol. 253, no. 3, pp. 374–402, 1986.
[236] L. C. Katz and E. M. Callaway, "Development of local circuits in mammalian visual cortex," *Annu.Rev. Neurosci.*, vol. 15, pp. 31–56, 1992.
[237] P. A. Salin and J. Bullier, "Corticocortical connections in the visual system: Structure and function," *Physiol. Rev.*, vol. 75, no. 1, pp. 107–154, 1995.
[238] R. R. Johnson and A. Burkhalter, "Microcircuitry of forward and feedback connections within rat visual cortex," *J. Comp. Neurol.*, vol. 368, no. 3, pp. 383–398, 1996.
[239] T. J. Buschman and E. K. Miller, "Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices," *Science*, vol. 315, pp. 1860–1862, 2007.
[240] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vis. Res.*, vol. 40, no. 10–12, pp. 1489–1506, 2000.
[241] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Rev. Neurosci.*, vol. 2, pp. 194–203, 2001.
[242] *Neurobiology of Attention*, L. Itti, G. Rees, and J. K. Tsotsos, Eds. Burlington, MA: Elsevier, 2005.
[243] S. Grossberg, "How hallucinations may arise from brain mechanisms of learning, attention, and volition," *J. Int. Neuropsychol. Soc.*, vol. 6, pp. 579–588, 2000.
[244] S. Grossberg and R. Raizada, "Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex," *Vis. Res.*, vol. 40, pp. 1413–1432, 2000.
[245] M. Luciw and J. Weng, "Where what network 4: The effect of multiple internal areas," in *Proc. IEEE 9th Int. Conf. Develop. Learn.*, Ann Arbor, MI, Aug. 18–21, 2010, pp. 311–316.
[246] J. Weng, "A general purpose brain model for developmental robots: The spatial brain for any temporal lengths," in *Proc. Workshop Bioinspired Self-Organizing Robot. Syst., IEEE Int. Conf. Robot. Autom.*, Anchorage, AK, May 3–8, 2010, pp. 1–6.
[247] G. Bi and M. Poo, "Synaptic modification by correlated activity: Hebb's postulate revisited," *Annu.Rev. Neurosci.*, vol. 24, pp. 139–166, 2001.
[248] Y. Dan and M. Poo, "Spike timing-dependent plasticity: From synapses to perception," *Physiol. Rev.*, vol. 86, pp. 1033–1048, 2006.
[249] Y. Wang, X. Wu, and J. Weng, "Synapse maintenance in the where-what network," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, Jul. 31–Aug. 5 2011, pp. 1–8.
[250] , I. Stockman, Ed., *Movement and Action in Learning and Development: Clinical Implications for Pervasive Developmental Disorders*. San Diego, CA: Elsevier, 2004.
[251] B. Z. Mahon, S. C. Milleville, G. A. Negri, R. I. Rumiati, A. Caramazza, and A. Martin, "Action-related properties shape object representations in the ventral stream," *Neuron*, vol. 55, no. 3, pp. 507–520, 2007.
[252] C. Yu, L. B. Smith, H. Shen, A. F. Pereira, and T. Smith, "Active information selection: Visual attention through hands," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 2, pp. 141–151, Aug. 2009.
[253] R. Held and A. Hein, "Movement-produced stimulation and the of visually guided behaviors," *J. Comp. Physiol. Psychol.*, vol. 56, pp. 872–876, 1963.
[254] R. C. Arkin, *Behavior-Based Robotics*. Cambridge, MA: MIT Press, 1998.
[255] D. B. Lenat, G. Miller, and T. T. Yokoi, "CYC, WordNet, and EDR: Critiques and responses," *Commun. ACM*, vol. 38, no. 11, pp. 45–48, 1995.
[256] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation for autonomous mental development," *IEEE Trans. Evol. Comput.*, vol. 11, no. 2, pp. 265–286, 2007.
[257] X. Huang and J. Weng, "Inherent value systems for autonomous mental development," *Int. J. Human. Robot.*, vol. 4, no. 2, pp. 407–433, 2007.
[258] J. L. Krichmar and G. M. Edelman, "Brain-based devices for the study of nervous systems and the development of intelligent machines," *Artif. Life*, vol. 11, pp. 63–77, 2005.
[259] J. L. Krichmar, "The neuromodulatory system: A framework for survival and adaptive behavior in a challenging world," *Adapt. Behav.*, vol. 16, no. 6, pp. 385–399, 2008.

Juyang Weng (S'85–M'88–SM'05–F'09) received the B.Sc. degree in computer science from Fudan University, Shanghai, China, in 1982, and the M.Sc. and Ph.D. degrees in computer science from the University of Illinois at Urbana-Champaign, in 1985 and 1989, respectively.

He is currently a Professor of Computer Science and Engineering at Michigan State University, East Lansing. He is also a Faculty Member of the Cognitive Science Program and the Neuroscience Program at Michigan State University. Since the work of Cresceptron (ICCV 1993), he expanded his research interests in biologically inspired systems, especially the autonomous development of a variety of mental capabilities by robots and animals, including perception, cognition, behaviors, motivation, and abstract reasoning skills. He has published over 250 research articles on related subjects, including task muddiness, intelligence metrics, mental architectures, vision, audition, touch, attention, recognition, autonomous navigation, and other emergent behaviors.

Dr. Weng is an Editor-in-Chief of the *International Journal of Humanoid Robotics* and an Associate Editor of the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, as well as a member of the Executive Board of International Neural Network Society. He was a Program Chairman of the NSF/DARPA funded Workshop on Development and Learning 2000 (1st ICDL), the Chairman of the Governing Board of the International Conferences on Development and Learning (ICDL) (20052007), Chairman of the Autonomous Mental Development Technical Committee of the IEEE Computational Intelligence Society (20042005), a Program Chairman of 2nd ICDL, a General Chairman of 7th ICDL (2008) and 8th ICDL (2009), an Associate Editor of the IEEE TRANSACTIONS ON PATTERN RECOGNITION AND MACHINE INTELLIGENCE, and an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He and his co-workers developed SAIL and Dav robots as research platforms for autonomous development.