2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

A Bridge-Islands Model for Brains: Developing Numeric Circuits for Logic and Motivation

Juyang Weng
Department of Computer Science and Engineering
Cognitive Science Program, and Neuroscience Program
Michigan State University
East Lansing, Michigan, 48824, USA
Email: http://www.cse.msu.edu/˜weng/

***Abstract*—Neuroscience has made impressive advances, but there is a lack of an overall computational brain theory. I would like to present a simplified computational theory in an intuitive language about how the brain wires itself as a multi-interchange bridge that bi-directionally connects many islands where each island is a sensor or effector. The wiring process of the brain is highly self-supervised while a baby lives and acts in his physical environment, e.g., sucking a milk bottle. I use a new precise framework of emergent finite automata to explain how the brain develops its numeric circuits that can be clearly understood in terms of logic. I also explain how the self-wired basic circuits become motivated through four additional neural transmitters beyond glutamate and GABA — serotonin, dopamine, acetylcholine, and norepinephrine. I use finite automata for precise and rigorous analysis and optimality.**

I. Introduction

Let us borrow a well-known tale originated from the Indian subcontinent, known as *Blind Men and an Elephant*. The blind men cannot visually see an elephant so each of them touches a body part of the elephant, but only one body part. None of them can tell correctly what an elephant is like. The brain is like an elephant; and each expert in one of the six disciplines — biology, neuroscience, cognitive science, computer science, electrical engineering, and mathematics — is like a blind man, as far as the brain is concerned. The existing partition of scientific disciplines is a major reason for this discipline-wide lack of necessary knowledge of other disciplines. E.g., it is well known that neuroscience is data rich and theory poor but many computer scientists were not motivated by brain challenges.

The term brain here is used as an intuitive term for the Central Nervous System (CNS), which includes both the spinal cord and the brain proper. Thus, we intuitively assume that the skull encloses the CNS. Anything inside the skull is internal; anything outside it is external.

A. Symbolic school

Two types of models have been used to model cognitive architectures, symbolic and emergent. The term "emergent" as proposed in Weng 2012 [31] is related to, but stricter than, a commonly used term "connectionist". Many connectionist representations are not fully emergent — do not emerge autonomously inside the skull-closed brain through development. They fall into the symbolic category, since many symbolic models also use connections. Except for trivial cases, there are no systems that do not need any representations.

To use a symbolic type, one assumes that there is a one-to-one correspondence between every hand-selected symbol and its meanings in a specified domain — each symbol has only a fixed set of meanings and each meaning has only one or a fixed set of symbols (e.g., Minsky 1991 [19], Anderson 1993 [3], Rosenbloom, Laird & Newell 1993 [20], Yuille & coworkers 2005 [26]), including some that are meant to model cortex: Albus 1991 [1], Lee & Mumford 2003 [17], George & Hawkins 2009 [9], Taylor et al. 2009 [24], Grossberg and coworkers 2009 [7], Albus 2010 [2], and Tenenbaum et al. 2011 [25]. Because of the symbol use, the modelers use "open-skull" approaches:

An open-skull approach requires that the skull is open during (manual) development of a mind. The holistically-aware central controller is the human designer outside the open skull. He defines each internal entity (e.g., a neuron or a module) using a symbolic meaning but its abstraction is an illusion, only in his mind and probably also in other people who hear his explanation. Each internal entity itself does not have a power of autonomous abstraction from real-world instances because of human *manually* and *internally* administrated symbolic training (e.g., using Bayesian framework).

As discussed below, such a symbolic model does not do task-nonspecific autonomous development.

B. Emergent school

To use an emergent type, one draws inspiration from brain-like emergent internal representations. As we will see later, our definition for emergent representations does not allow a human programmer to handcraft the contents or boundaries of extra-body concepts, since representations in this category use a "closed-skull" approach:

A closed-skull approach requires that the skull is closed during (autonomous) development of a mind, from the inception time throughout the life. Only the brain's sensory ends S and the motor ends M are open to the external environment).

Many artificial neural networks satisfy this "closed-skull" definition. Examples include the unsupervised Self-Organizing Maps by Kohonen 1982 [15], feed-forward computation networks with error back-propagation learning [38], [6], [16], [22] (which can incrementally learn but either do not have

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

a long-term memory or have a rigidly frozen old memory which does not share resource with new experience); Hopfield network [13] (which converges into a local attractor from an initial guess); ARTMAP by Carpenter et al. 1991 [4] (which finds the nearest bottom-up match using top-down consecutive selection); Cresceptron by Weng et al. 1992 [33], [34] (which incrementally develops a feed-forward network which learns local-to-global features with increasing location tolerance from early to later layers); Elman network 1993 [5] (which has a local recurrent layer, for temporal context, using error back-propagation learning); the recurrent LISSOM network by Sit & Miikkulainen 2006 [23]; the feedforward networks by Serre et al. 2007 [21]; the Deep Learning Networks by Hinton 2007 [11]; and the MILN network by Weng et al. 2008 [35]. An emergent representation for extra-body concepts emerges autonomously from interactions with the internal environment and the external environment.

This characterization of emergent model is a refined, stricter definition compared with the traditional type called *connectionist models*, which traditionally means distributed representations in a network. However, the term "connectionist" is misleading and imprecise. For example, the symbolic base network of Hidden Markov Models (HMM) is also a network. A probability distribution inside a symbolic base network could be mistaken as a connectionist model. I will use the term "emergent model" below instead.

An agent is anything that senses and acts. "Emergent" in "emergent model" means: (a) The internal representations, inside the closed skull, emerge while the brain-like network interacts with the *external environment* through its sensor ports and effector ports that are the only input-and-output ports open to the external environment. (b) The network adaptation and operation is regulated by the Developmental Program (DP) whose programmer could not precisely predict the contents of the lifetime extra-body environments, such as the nature of the tasks and the concepts required. For a natural agent (e.g., animal), the DP is the genome, whose programmer consists of all the predecessors and the nature met by the predecessors. For an artificial agent (e.g., robot), the DP is a computer program, whose programmer is a human being, but he has left the agent after the agent's "birth".

Because an emergent representation is only indirectly related to symbolic meanings (e.g., a 67° edge), the network learning mechanisms are not specific to particular meanings either. This important property is useful for understanding development, as the set of learning mechanisms must work through different ages, through which the agent (natural or artificial) learns an open number of skills and performs an open number of tasks. Often, we use some symbolic meanings as examples (e.g., a neuron in V1 detects a 67° edge) to assist understanding of the functions of emergent representations, but such explanations are incorrect and misleading.

C. Take the best of the two schools

The brain model here takes the best of the above two schools: an emergent model that can abstract logically, as discussed in [29].

In the remainder of this paper, I first present an overview of the brain architecture that is novel and very different from the other models about the brain in that they all explicitly model the Brodmann areas, instead of enabling them to emerge by this model depending on the sensory and motor signals. Such prior models are inconsistent to our understanding that in the born blind, visual cortex is nonexistent and the cortical resource is reassigned to audition and touch (Voss 2013 [27]). Therefore, the model here seems to be the only work I am aware of that is consistent with the above known cross-modality plasticity of the brain.

II. The Bridge-Islands Model of the Brain

Through the course of evolution, the genome of each species has evolved to better regulate the development of each individual in order to survive and compete successfully in the species' own environment. For example, the environment of a mouse is very different from that of a human.

The brain-mind of an adult mouse is very different from that of an adult human, but the developmental mechanisms of mammalian brains seem to be very similar. Let us consider mainly humans here.

The development of each individual starts from a single cell — a fertilized egg called a zygote. The zygote gradually develops into a newborn and further into an adult. This process involves two parallel and inter-dependent processes, the body development and the brain-and-mind development.

Cells split into more and more cells. The cells migrate, differentiate, and connect, through interactions with other cells in each cell's environment. When cells turn into different types and contact, they form body parts, such as tissues, organs, and circuits. The way the brain wires itself depends on the activities of the sensors and effectors. This is true not only after birth, but also before birth.

The brain is not only a signal processor, but first and more importantly also the developer of the signal processor. A static diagram of the brain circuit, no matter how complete it is, is far from enabling us to understand how the brain works. Without feeling the current "traffic", the brain "roads" cannot expand.

A human has mainly five senses — sight, hearing, taste, smell, and touch. Each sense has a sensor, e.g., eye for sight and ear for hearing. A human has many effectors for carrying out actions. There are two types of effectors — muscles and glands.

From a fetus to infancy to adulthood, the brain constantly wires and rewires itself to handle two-way joint communications among all the sensors and all the effectors.

Let us use an analogy: Suppose that each sensor is an island. Each effector, such as (the muscles in) a body part or a gland, is also an island. Then, the body has many sensory islands and many effector islands. The brain is a multi-exchange bridge that autonomously builds itself to connect *from and to* all the islands, as illustrated in Fig. 1

Before birth, all the sensors and effectors already have spontaneous signals that guide the brain's prenatal wiring. The retina has spontaneous waves of signals although the baby's eyes are still closed. The baby's legs kick in the mother's womb.

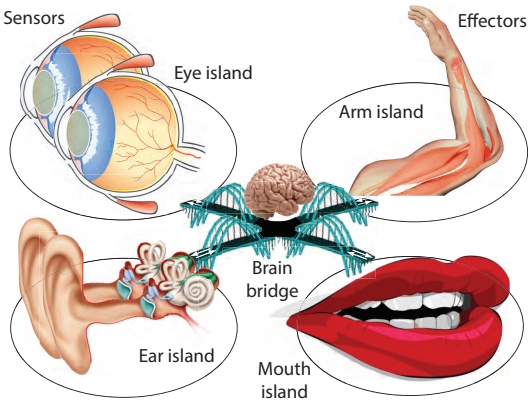

Fig. 1. An illustration of the bridge-islands brain model. The brain is a multi-exchange super bridge that bi-directionally connects with all islands. The bridge produces winner neurons that form an ever-changing committee that collectively vote to report the combined configuration of the firing patterns of all the current islands. Each island uses the votes from the committee to predict the next firing pattern in the island. There are two types of islands, sensory and effector. Each sensory island is supervised by the external environment and, therefore, its firing pattern is always concrete and cluttered with many objects' signals. Each effector island is either supervised by the external environments or self-emerging and, therefore, its firing patterns can be abstract, representing values of an abstract concept. Each effector island is a hub of attention. Its firing pattern enables only the voting members that correspond to the attended sensory part to win, fire and vote. Then, the brain attends that sensory part at that time instant.

Such spontaneous activities allow the prenatal brain to form coarse wiring for estimating coarse statistics across all of its sensory islands and effector islands, as well as generating inborn reflexes. The heart beats well before birth and the baby cries right after birth.

Reported by many neural anatomy studies [8], the brain circuits use more than local recurrence (e.g., Hopfield networks), do not use random connections (called reservoir networks), and have more than a cascade of processors (called deep learning networks) between any pair of sensor and effector.

The brain does not seem to find errors in the effector ends and then do error back-propagation. The baby does not "know" the error. Asking all the connected neurons to change their weights to reduce the current errors causes a loss of long-term memory.

The brain's self-organization does not have an internal "government" — a central controller. When a brain seems to autonomously control its actions, it only responds to the environments, inside and outside the skull.

There exists no standard symbol for a concept. When I act (e.g., say "apple") every time the muscle signals are different. I may use a slightly different tone, volume, or speed. Our model allows variation in each action.

The brain needs to attend to individual objects in a cluttered scene. When you pick an apple from an apple tree, your action depends on only the "pixels" of that apple but not many other "pixels" from other parts of the tree. Thus, each neuron for vision must have a limited receptive field in the retina. It responds to only some area in the retina, not the entire retina.

Further, each brain neuron responds to not only sensors but also effectors. Effectors too? Yes, patterns of effector firing are concept states, as we see below.

What does the brain do? Each of its feature neurons represents a set of similar joint firing patterns of all the islands so that each island can use the firing feature neurons to predict its own next firing pattern. Intuitively, all the feature neurons inside the brain are like tiles of different sizes and shapes called Voronoi regions that together seamlessly tile the "floor" of all the islands.

Suppose that only the best matched neuron, or "winner neuron" in the bridge fires. All neurons in each island that fire next are linked from this firing bridge neuron. Thus, a single bridge neuron is able to trigger the correct next firing pattern in every island. In general, more winner bridge neurons fire, they correspond to a dynamic committee that "votes" for, through links, all the next firing neurons in every island.

Each bridge neuron has a *sensory receptive field* (e.g., from the apple pixels), an *effective receptive field* (e.g., from the muscles for saying "apple" and for the arm reaching for the apple), and a *lateral receptive field* (from other bridge neurons).

Since there is no central controller, each neuron does not "know" its brain's role when it is generated from mitosis.

The brain has two large categories of neurons [14] — projection neurons and interneurons. Projection neurons are feature detectors (for all the above three input sources) that can connect far. Interneurons connect only near. Each interneuron turns input signal from "positive" (excitatory via Glutamate transmitters) to "negative" (inhibitory via GABA transmitters). All projection neurons inhibit other projection neurons through the help of many interneurons.

The net effect of the mutual inhibitive competition is that far fewer winner neurons can fire at any time. In the remainder of this paper, by "neuron", we mean a projection neuron helped by interneurons.

Each neuron gets its pre-action membrane potential to fire from three sources of input — bottom-up from sensors, lateral from other neurons, and top-down from effectors. It has a set of synaptic weights (vector, i.e., many signal lines), as its long-term memory, to match each source of input (vector). The better the match between the synaptic weights with the input source, the more pre-action potential the neuron gets.

Through lateral competition, only the top-winner neurons that have a high pre-action potential can win to fire. Thus, for a neuron to consistently win in competition within a short time window, it must match well not only the bottom-up source (e.g., an "apple" image patch at a retinal location), but also top-down source (e.g., muscles saying "apple" or muscles for the arm holding the "apple", or best both sources). Thus, the firing of each neuron is also affected by your actions (i.e., your intentions in general).

A neuron updates its weights only when it wins and fires, which means that the current sensory input and effect input is its "business". Firing neurons at any time are the current working memory, while all other neurons that do not match well and do not fire are the current long-term memory.

When a neuron fires, all its synapses (bottom-up, lateral, and top-down) update their conductance (weights) incremen-

tally so that every conductance statistically matches the input better in the future. This is qualitatively known as the Hebb's rule — neurons that fire together wire together [14]. The theoretical quantitative Hebb's rule is statistically optimal [29].

From early embryo through adulthood, proteins and other molecules called *morphogens* released from other cells guide the migration and connection of every new cell in a coarse-to-fine manner [14] . Because of this manner, every two mature neurons, if their firings are statistically strongly correlated or anti-correlated, tend to connect with each other, excitatorily or inhibitorily. Otherwise, their connections retract, if their connections existed previously.

How does a baby constantly self-supervise during his daily activities?

Suppose that he is sucking milk from a milk bottle. When he was born, he did not know what milk bottles were. His sucking effector "tells" the type of the object — milk bottle, since he cannot say "milk" yet. His arm that holds the milk bottle "tells" the relative location of the object on the retina.

Statistically, only those bridge neurons that match well for both bottom-up input and top-down input can survive the competition to consistently fire. The bottom-up energy corresponds to the match of the milk-bottle image patch in correct location and pattern, but not other locations or another pattern at this location. The top-down energy corresponds to the match of the sucking effector and the arm effectors, but not non-sucking effectors and other arm positions.

Thus, the "suck" neurons in the "vocal effector island" co-fire with, and then link with, all the bridge neurons for different locations and appearances of the milk bottle. They become specific concept neurons for type *milk bottle*, invariant to locations and appearances [28].

Likewise, the "location" neurons in the "location effector island" co-fires with, and then links with, all the bridge neurons for different object types and appearances at that location. They become specific concept neurons for *that particular location*, invariant to object types and appearances, milk bottle, his palm, or any other objects [28].

Many modelers thought that the state of a brain must be totally inside the skull. It is interesting to note that firing patterns of all effector neurons become the brain's concept states. Every *learnable* action (e.g., muscles, not glands) of the brain is observable, teachable, and calibratable by the self and teachers. When you think, you "tacitly" do muscle concept rehearsal. Such concept states become more adult-like knowledge while the child grows up, beyond "type" and "location", to include any concept a human can be taught. Namely, muscle states are intentions learned from experience.

As the Developmental Network (DN) theory [29], [30] has formally proved, the brain is an emergent finite automaton. A symbolic finite automaton can only be handcrafted but the brain's finite automaton autonomously emerges.

The set of next firing bridge neurons depends on not only the sensory image (e.g., many apples on an apple tree) but also effector state (e.g., apple in the type effector instead of leaves; this location in the location effector instead of other locations.)

Although such emergent networks learn invariances, they do not have any motivation. The brain uses different types of neural transmitters to modulate its network.

The serotonin transmitters released by a pain signal suppress the firing of effecter neurons since they are likely responsible for the punishment [37]. The dopamine transmitters arising from a candy reward excite the firing of the effecter neurons because they are likely responsible for the reward [37]. Both serotonin and dopamine increase the learning rate of the firing bridge neurons so that such important events are learned faster and harder to forget [40]. The serotonin and dopamine systems give rise to pain-avoidance and pleasure-seeking motivation.

Some researchers argued that acetylcholine and norepinephrine transmitters signal *expected* and *unexpected* uncertainties, respectively [39]. The acetylcholine transmitters from each neuron signal the deviation between inputs and the synaptic weights across the neuron. The norepinephrine transmitters released from each synapse indicate the deviation between the pre-synaptic input and the synaptic conductance. The ratio of norepinephrine over acetylcholine indicates the novelty at each synapse. This novelty dynamically regulates the growth or retraction of each synapse, to enable each neuron to dynamically search for the domain of most common input patterns. Therefore, each neuron can detect the shape of the object it was assigned to detect by competition [28].

The glutamate and GABA transmitters for the basic brain circuits, plus the above four types of neural transmitter, seem to form a minimum set of neural transmitters necessary for motivating the brain. All the brain areas emerge autonomously, including the networked internal hierarchy [8], as a consequence of statistics from the activities of life, instead of being fully specified by the genome or statically handcrafted by a human programmer. This developmental model of the brain is also consistent with why, in the brains of the blind at birth, the visual areas are assigned to audition and touch. In contrast, a static brain pathway model cannot explain such assignments.

Next, I will use the well known automata theory to explain that the above theory is not only intuitive, but is also built on precise and clearly understandable mathematical logic. First, we must discuss the types of representation.

III. Analysis

In this section, I analyze how the above bridge-islands model can perform both types of skills, declarative and non-declarative skills. They are all performed by muscles, where each action in an island corresponds to a sequence of firing patterns in each island.

The *internal representations* of an agent refers to the ways in which information is organized inside the "brain" of the agent, natural or artificial.

The architecture of an agent does not necessarily specify actual representation inside the agent's brain, although the agent architecture may affect the representations. Therefore, we will use well defined representation types as a central issue in our discussion. All the internal representations fall into two large categories, symbolic and emergent, illustrated in Fig. 2 and defined as follows:

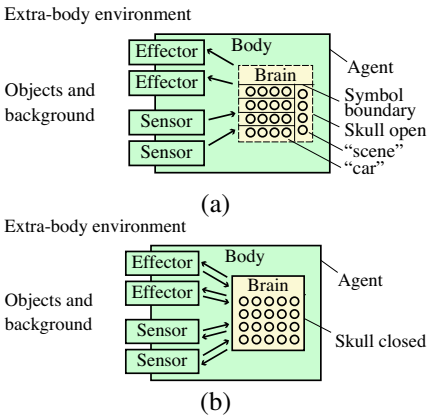

Fig. 2. Agents using (a) symbolic representation and (b) emergent representation where a brain-bridge bi-directionally connects with two types of islands, sensory islands and effector islands. In (a) the skull is open during manual development. This is an external model as the human programmer only represents brain's external behaviors. The human programmer handpicks a set of task-specific concepts using human-understood symbols and he handcrafts the boundaries that separate concepts. In (b) the skull is closed during autonomous development. The human programmers only design task-nonspecific mechanisms for autonomous development. The internal representations autonomously emerge from experience. Many neural networks do not use the feedbacks connections in (b).

Definition 1 (Symbolic and emergent): A symbolic representation in the brain of an agent has human *handcrafted* contents and boundaries where each zone represents a *symbol* (e.g., text as label) about a concept of the *extra-body* environment (e.g., car). An emergent representation does not allow such human handcrafted (or genome fully specified) contents or boundaries.

There are always new extra-body concepts that the parents' genome does not "know" about (e.g., Internet). A symbolic representation cannot explain how the brain learns such new concepts. There is no rigid boundary in the brain that prevents the representation for an extra-body concept to penetrate into that for another. For instance, an edge orientation to be detected can be associated with any other sensory event and motor event, depending on the task to be learned. Internally, the representation of this edge orientation penetrates into the representations for those associated events, and verse versa, due to the two-way connections, direct or indirect, among the co-firing neurons. For more discussion and evidence of brain support, see Weng 2011 [31].

A. Finite automata

As Harnad [10] and many others argued, how does the brain network do arbitrary re-combinations like a computer on the fly? The FA theory is a model, although it is symbolic.

The classical definition of FA is for a language acceptor. We must extend the definition to be useful for brain-mind:

Definition 2 (agent FA): An agent FA (AFA) M for a finite symbolic world is a 4-tuple $M = (Q, \Sigma, q_0, \delta)$, where Σ is the set of input symbols (alphabet), Q is the set of states but also output symbols, q_0 is the starting state, $\delta : Q \times \Sigma \mapsto Q$ is the state transition function.

The meanings of each $\sigma \in \Sigma$ and each $q \in Q$ are not part of the AFA. Like all symbolic models, the meanings are in the mind of human designers. Since human designers understand human languages, the meanings are often explained in a human language.

Fig. 3 gives an example of AFA. It has meanings expressed in English, where we simply use natural language text to replace every $\sigma \in \Sigma$. After reading the meanings, a human understands that an AFA can be very general, simulating an animal in a micro, symbolic world. It is important to note that for each state in AFA, the cognition set can also be considered part of the action set, as a cognitive state can always correspond to a report action.

Let Σ^* denote the set of all possible strings of any finite number of symbols from Σ. It has been proved [12], [18] that an FA with n states with alphabet Σ partitions all the strings in Σ into n sets, each set corresponding to a particular state. The set of all strings that lead to state q is denoted as $[q]$. Thus, we have

$$\Sigma^* = [q_0] \cup [q_1] \cup ... \cup [q_n].$$

with $[q_i] \cap [q_j] = \phi$ for all $i \neq j$,

From q_0, strings in $[q]$ all arrive at the same state q. That is, all these strings are indistinguishable by the AFA from this point on. We define a relation between any two strings. String α and β are related denoted as $\alpha R \beta$ if and only if all the possible strings $\gamma \in \Sigma^*$, the two concatenated strings $\alpha\gamma$ and $\alpha\gamma$ lead to the exactly the same symbolic output. Suppose α, β and γ are any three strings in Σ^*. By definition, an equivalence relation has three properties, reflexivity ($\alpha R \alpha$), symmetry (if $\alpha R \beta$ then $\beta R \alpha$) and transitivity (if $\alpha R \beta$ and $\beta R \gamma$ then $\alpha R \gamma$). It has also been proved [12], [18] that this relation among strings $\alpha \in \Sigma^*$ as defined above is indeed an *equivalence relation* on Σ^*, having the above three properties.

This equivalence relation might cause a potential problem if we merge two strings into the same state but they should not be treated exactly the same in the future. For example, in Fig. 3 the set $[z_3]$ consists of all strings that end with "kitten" or "young cat" and followed by any number of "well". If "kitten" and "young cat" needs to be treated differently in the future, they should not arrive at the same state z_3. Once they fall into the same state, AFA no longer can distinguish their difference, since it does not have internal representations. But, DN has internal representations that can tell whether "kitten" or "young cat" is actually received.

B. Symbolic networks: Probabilistic variants

An FA has many probabilistic variants (PVs), e.g., HMM, POMDP, and Bayesian Nets. Like the FA, each node of a PV is defined by the handcrafted meaning which determines what data humans feed it during training. A pre-processor can expand symbolic inputs to real vector inputs (e.g., images) based on handcrafted features (e.g., Gabor filters). The PV determines a typically better boundary between two ambiguous symbolic nodes using probability estimates, e.g., better than the straight nearest neighbor rule. However, this better boundary does not change the symbolic nature of each node. The base network of any PV is still the symbolic FA. Therefore, FA and all its PVs are all called Symbolic Networks (SNs) here.

Since both input $\sigma \in \Sigma$ and output $q \in Q$ are external, an AFA does not have internal representations. Likewise,

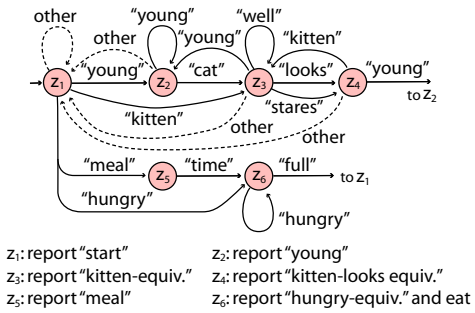

Fig. 3. Agent Finite Automaton (AFA). Output: symbol representing the current state (circle). Input: symbolic label for the arrow from the current state. A label 'other' means any symbol in Σ other than the symbols marked from the state. The AFA starts from state z_1. The meanings of symbols are not part of the AFA, only in the mind of human designers. An AFA does not "know" the meanings. The meanings of input symbols are indicated by an English word. The meanings of each state are expressed in English text below the AFA, always represented as a set of actions. For example, state z_4 means that the equivalent meaning of the attended last subsequence is "kitten looks" or equivalent. The "other" transitions from the lower part are omitted for graph clarity.

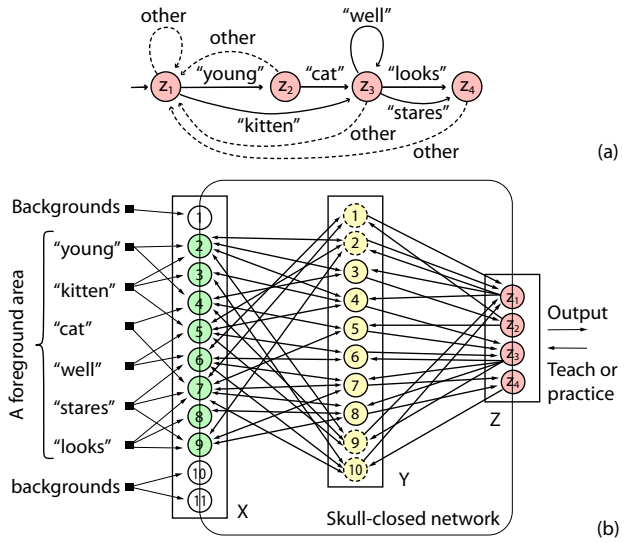

Fig. 4. Relate a "skull-open" AFA with a "skull-closed" DN. (a) An AFA, handcrafted and static, reasons in the symbolic world. (b) A corresponding DN as a bridge-islands model of the brain that lives and learns autonomously in the real physical world. The X area corresponds to all sensory islands. the Z area corresponds to all effector islands. Only one (green) foreground is shown, but the DN has many other Y neurons to deal with all possible locations of the (green) foreground (not shown for graph clarity). That is, the object images like "young", "kitten", can appear anywhere in the retina with natural scales and appearances. The DN in (b) can be taught to produce the same equivalent actions as the AFA in (a). An AFA does not have any internal representation but a DN has (inside the skull). An AFA is not grounded — inputting a symbol in Σ and outputting a symbol in Q at a time. However, a DN is grounded — inputting-and-outputting $\mathbf{x} \in X$ and $\mathbf{z} \in Z$ at a time, where $\mathbf{x}$ contains irrelevant components (e.g., other objects and backgrounds). The DN does attention but the AFA does not. In (b) the "dashed" cells in Y correspond to the "dashed" transitions in (a).

any SN does not have any internal representation either, as its probability distributions are about external symbols. It is the author's view that all SNs only model brain's external behaviors. All the meanings of $\sigma \in \Sigma$ and $q \in Q$ are external, only in the eyes of the human designers. The SN does not "know" the meanings of each symbol.

C. Autonomous mental development

The paradigm of Autonomous Mental Development (AMD) does not model a static adult brain directly. Instead, it considers how the body and the brain develops from conception, through fetus, infancy, and adulthood. Probably the two most striking hallmarks of AMD, different from traditional machine learning and traditional models for a mind, are:

1) task nonspecificity [36],
2) the skull is closed throughout development, and
3) all representations *emerge* from interactions with the external environment and the internal environment.

Task non-specificity of development means: (a) For natural brain-mind, evolution cannot predict the exact task and the exact task environment at any particular lifetime, nor can a result of the evolution — the DP (genome). As part of the phenotype, inborn behaviors are part of the body, mainly about the intra-body concepts (e.g., reflexes), not much about the extra-body concepts, due to very limited exposure to the extra-body environment (e.g., outside the womb). (b) For artificial brain-mind, no task is given during the DP programming time. Because of the above three hallmarks, a developmental agent should not use any symbolic representation. Even the task is unknown, let alone extra-body concepts. Symbols that represent a static set of handcrafted extra-body concepts seems impossible for the developmental brain-mind.

D. Developmental Network (DN)

Let us consider a basic network, called Developmental Network (DN). Because any model about the nature is always an approximation, I hope that DN can explain basic principles about the brain-mind of Eumetazoa — a clade within the animal group excluding a few exceptions (e.g., sponges) that do not have nerves. They are multicellular, eukaryotic organisms in the kingdom animalia, including simple ones (e.g., a fruit fly) and complex ones (e.g., a human). Of course, there are many biological differences across these species, e.g., receptor types and neuron types.

A DN has three areas X, Y, and Z. For notation simplicity, we use such a letter (e.g., X) to denote the physical area as well as the space of all possible response patterns of the area.

Exemplified in Fig. 4(b), a DN serves as a model of the brain-mind which lives and learns autonomously in the open-ended, dynamic, real physical world. In general a DN has three areas, the internal area Y as a "bridge", its sensory area X as a "island" relative to the Y and its motor area Z as another "island" relative to Y.

A DN can serve as a model for the entire brain. If the island X consists of all sensors (e.g., retina, cochlea, skin) of the brain and the island Z consists of all effectors (e.g., muscles and glands) of the brain, the DN is a model for the entire brain-mind.

A DN can also serve as a model for any internal brain area (e.g., a Brodmann area, a union of Brodmann areas, where the internal area Y is a bridge and its two connected areas X and Z are islands of the bridge, all inside the brain.

For convenience of discussion, the default ascending order of the three areas are X, Y, and Z. However, such an ascending order is not always clear, e.g., LIP which links two streams, the ventral and dorsal streams. In the DN algorithm, the roles of X and Z are symmetrical. The world can supervise both X and Z, directly or indirectly.

I predict that the most basic function of any brain area as a bridge is to predict: Take the current input $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, where $\mathbf{x} \in X$, $\mathbf{y} \in Y$, and $\mathbf{z} \in Z$, and use the adaptive part N of the DN, to predict what is next as $(\mathbf{x}', \mathbf{y}', \mathbf{z}')$ and update the adaptive part to N':

$$(\mathbf{x}', \mathbf{y}', \mathbf{z}', N') = f_{\rm DN}(\mathbf{x}, \mathbf{y}, \mathbf{z}, N) \quad (1)$$

where $f_{\rm DN}$ denotes the predictive function of DN.

To accomplish that, each area A of X, Y, and Z learn and computes using the same framework for the area function. The area Y takes X and Z as input islands. Similarly, each island, X and Z, takes Y as input. To simulate any given AFA, the response of the Y area is sparse but that in X and Z can be any practical. Each neuron in A is a feature cluster/detector. The Y area uses the input data from (X, Z) to initialize its neuronal clusters. Each time, only top-k (e.g., $k = 4$) neurons that best match the current (X, Z) input win, fire and update. The update rule is such that Y best represents (X, Z) space in the sense of maximum likelihood. Consider $k = 1$ for simplicity. Then, using Hebbian learning, all the firing neurons in each island links from the firing Y neuron. Thus, the Y neuron always predicts only the correct firing neurons in each island. I have proved that for an FA with n transitions, a DN with n Y neurons can perfectly learn the input-output pair of FA incrementally, immediately, and error-free [32]. When the input set is infinite, the DN is further optimal in the sense of maximum likelihood [32]. For the detail of DN, the reader is referred to [29].

The following gave two examples about how the GDN learns an AFA.

Example 1 (AFA as a number tracker): GDN simulates the AFA in Table I, which is a number tracker. The X area has two possible inputs $\Sigma = \{\sigma_1, \sigma_2\}$, and the Z area has four possible inputs $Q = \{q_1, q_2, q_3, q_4\}$. Each input σ_2 makes the GDN increment in a loop way through the four states. Each input σ_1 makes the GDN to round to the nearest extremum, 1 for q_1 or 4 for q_4. Starting state: q_1.

The following table gives the state transition function $q' = \delta(q, \sigma)$, where the entry at row q and column σ, is the target state q':

TABLE I. STATE TRANSITION TABLE

δ	σ_1	σ_2
q_1	q_1	q_2
q_2	q_1	q_3
q_3	q_4	q_4
q_4	q_4	q_1

Table II shows how the teacher picks up a particular training sequence to run AFA while he teaches the GDN. For clarity, the table uses the time unit of network update. The row $Y(t)$ indicates the meaning of $\mathbf{y}(t) = (\mathbf{x}(t-1), \mathbf{z}(t-1))$. The short sequence shown has not included all the AFA state transitions. So, the learning of AFA has not been complete. From the rows $X(t)$:su and $Y(t)$: su, the reader can see if he can independently fill all the other rows based on the theory above.

The following is another example, which shows how a spatiotemporal GDN deals with pattern recognition where time is not necessary for the task.

Example 2 (AFA for pattern recognition): The teacher wants to teach a GDN to recognize two foreground objects α and β. Since a GDN is a spatiotemporal, the teacher uses a background image B to serve as a break between two objects so that a supervised action corresponds to the object, not the concatenation of two objects. The corresponding AFA let its state q' to classify the input σ. Although $q' = \delta(q, \sigma)$, the δ function is independent of q for this special AFA.

Table III shows how the teacher trains the spatiotemporal GDN for a spatial problem only and how the GDN learns. Although the task does not need time for a task-specific pattern recognition machine, the GDN is not a special-purpose single-frame based recognizer. As the task must be learned in real time from the real dynamic physical world in a grounded way, the GDN spends internal Y neurons to learn transitions that are necessary for the GDN to be aware of dynamic aspects beyond narrowly defined single-frame task. For example, the Y area has neurons for detecting dynamic events like B to α, α to α, and α to B.

IV. PRACTICAL SYSTEMS

Where-What Networks (WWN-1 through WWN-8) are experimental embodiments of the bridge-islands model here. Their Z area is taught with two or more categories of concepts (e.g., where and what). For example, the type concept is invariant to location and the location is invariant to type.

V. CONCLUSIONS

The bridge-island model treats the brain as a motivated self-emergent numeric circuit that generates and updates from experience. The effector islands are typically abstract or purposive, as hubs of sensory attention for sensory islands which are concrete and cluttered with many objects. Lower and higher intelligences are all predictions for islands. With more experience the sensor-effector skills become more complex.

REFERENCES

[1] J. S. Albus, "Outline for a theory of intelligence," *IEEE Trans. Systems, Man and Cybernetics*, vol. 21, no. 3, pp. 473–509, May/June 1991.

[2] ——, "A model of computation and representation in the brain," *Information Science*, vol. 180, no. 9, pp. 1519–1554, 2010.

[3] J. R. Anderson, *Rules of the Mind*. Mahwah, New Jersey: Lawrence Erlbaum, 1993.

[4] 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 Networks*, vol. 4, pp. 565–588, 1991.

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

[6] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-CS-90-100, Feb. 1990.

TABLE II. THE DN TIME SEQUENCE TABLE FOR EXAMPLE 1. "EM": EMERGENT IF NOT SUPERVISED; "SU:" SUPERVISED BY THE ENVIRONMENT. "*" MEANS FREE. "-" MEANS NOT APPLICABLE. "?" MEANS A NEW Y NEURON FIRING THAT DOES NOT HAVE ANY LINK WITH AREAS X AND Z.

Time t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$Z(t)$: su	q_1	q_1	q_1	*	*	*	q_2	*	q_3	*	q_4	*	q_1	*	*	*	q_1	*	*
$Z(t)$: em	-	-	?	q_1	q_1	q_1	?	q_2	?	q_3	?	q_4	?	q_1	q_2	q_2	?	q_1	q_2
$Y(t)$: $\mathbf{z}$	-	q_1	q_1	q_1	q_1	q_1	q_1	q_2	q_2	q_3	q_3	q_4	q_4	q_1	q_1	q_2	q_2	q_1	q_1
$Y(t)$: $\mathbf{x}$	-	σ_1	σ_1	σ_1	σ_1	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_1	σ_1	σ_2	σ_2
$X(t)$: em	-	-	?	σ_1	σ_1	σ_1	?	σ_2	?	σ_2	?	σ_2	?	σ_2	σ_2	σ_2	?	σ_2	σ_2
$X(t)$: su	σ_1	σ_1	σ_1	σ_1	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_2	σ_1	σ_1	σ_2	σ_2	σ_2

TABLE III. THE DN TIME SEQUENCE TABLE FOR EXAMPLE 2. "EM": EMERGENT IF NOT SUPERVISED; "SU:" SUPERVISED BY THE ENVIRONMENT. "*" MEANS FREE. "-" MEANS NOT APPLICABLE. "?" MEANS A NEW Y NEURON FIRING THAT DOES NOT HAVE ANY LINK WITH AREAS X AND Z.

Time t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$Z(t)$: su	B	B	B	*	α	*	α	B	*	β	*	β	B	*	*	*	*	*	*	*
$Z(t)$: em	-	-	?	B	?	α	?	?	B	?	β	?	?	B	α	α	α	B	B	β
$Y(t)$: $\mathbf{z}$	-	B	B	B	B	α	α	α	B	B	β	β	β	B	B	α	α	α	B	B
$Y(t)$: $\mathbf{x}$	-	B	B	α	α	α	B	B	β	β	β	B	B	α	α	α	B	B	β	β
$X(t)$: em	-	-	?	α	?	α	?	?	β	?	β	?	?	α	α	α	B	β	β	β
$X(t)$: su	B	B	α	α	α	B	B	β	β	β	B	B	α	α	α	B	B	β	β	B

[7] 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," *Cognitive Psychology*, vol. 58, pp. 1–48, 2009.

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

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

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

[11] G. E. Hinton, "Learning multiple layers of representation," *Trends in Cognitive Science*, vol. 11, no. 10, pp. 428–434, 2007.

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

[13] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the USA*, vol. 79, no. 8, pp. 2554–2558, 1982.

[14] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, and A. J. Hudspeth, Eds., *Principles of Neural Science*, 5th ed. New York: McGraw-Hill, 2012.

[15] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

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

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

[18] J. C. Martin, *Introduction to Languages and the Theory of Computation*, 3rd ed. Boston, MA: McGraw Hill, 2003.

[19] M. Minsky, "Logical versus analogical or symbolic versus connectionist or neat versus scruffy," *AI Magazine*, vol. 12, no. 2, pp. 34–51, 1991.

[20] P. S. Rosenbloom, J. E. Laird, and A. Newell, Eds., *The Soar Papers*. Cambridge, Massachusetts: MIT Press, 1993.

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

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

[23] Y. F. Sit and R. Miikkulainen, "Self-organization of hierarchical visual maps with feedback connections," *Neurocomputing*, vol. 69, pp. 1309–1312, 2006.

[24] 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 and Visual Computing*, vol. 27, no. 11, pp. 1641–1657, 2009.

[25] 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.

[26] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," *Int'l J. of Computer Vision*, vol. 63, no. 2, pp. 113–140, 2005.

[27] P. Voss, "Sensitive and critical periods in visual sensory deprivation," *Frontiers in Psychology*, vol. 26, 2013, doi: 10.3389/fpsyg.2013.00664.

[28] Y. Wang, X. Wu, and J. Weng, "Skull-closed autonomous development: WWN-6 using natural video," in *Proc. Int'l Joint Conference on Neural Networks*, Brisbane, Australia, June 10-15, 2012, pp. +1–8.

[29] J. Weng, "Why have we passed "neural networks do not abstract well"?" *Natural Intelligence: the INNS Magazine*, vol. 1, no. 1, pp. 13–22, 2011.

[30] ——, *Natural and Artificial Intelligence: Introduction to Computational Brain-Mind*. Okemos, Michigan: BMI Press, 2012.

[31] ——, "Symbolic models and emergent models: A review," *IEEE Trans. Autonomous Mental Development*, vol. 4, no. 1, pp. 29–53, 2012.

[32] ——, "Establish the three theorems: DP optimally self-programs logics directly from physics," in *Proc. International Conference on Brain-Mind*, East Lansing, Michigan, July 27 - 28 2013, pp. +1–9.

[33] J. Weng, N. Ahuja, and T. S. Huang, "Cresceptron: a self-organizing neural network which grows adaptively," in *Proc. Int'l Joint Conference on Neural Networks*, vol. 1, Baltimore, Maryland, June 1992, pp. 576–581.

[34] ——, "Learning recognition and segmentation using the Cresceptron," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 109–143, Nov. 1997, cited early versions in IJCNN 1992 and ICCV 1993.

[35] J. Weng, T. Luwang, H. Lu, and X. Xue, "Multilayer in-place learning networks for modeling functional layers in the laminar cortex," *Neural Networks*, vol. 21, pp. 150–159, 2008.

[36] 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.

[37] J. Weng, S. Paslaski, J. Daly, C. VanDam, and J. Brown, "Modulation for emergent networks: Serotonin and dopamine," *Neural Networks*, vol. 41, pp. 225–239, 2013.

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

[39] A. J. Yu and P. Dayan, "Uncertainty, neuromodulation, and attention," *Neuron*, vol. 46, pp. 681–692, 2005.

[40] Z. Zheng, K. Qian, J. Weng, and Z. Zhang, "Modeling the effects of neuromodulation on internal brain areas: Serotonin and dopamine," in *Proc. International Joint Conference on Neural Networks*, Dallas, TX, 2013, pp. +1–8.