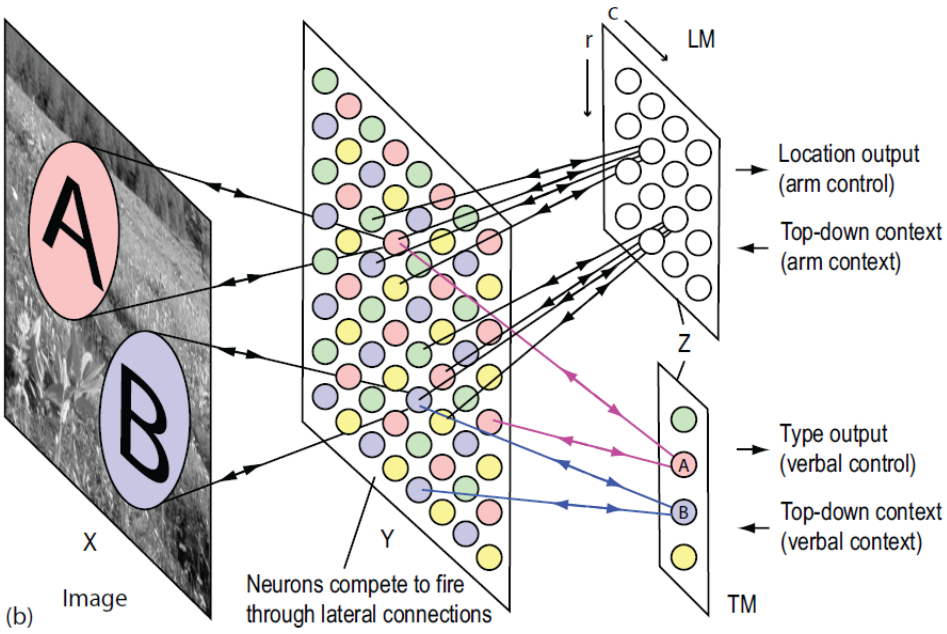
ISSN 2164-8522

Natural Intelligence

The INNS Magazine

Volume 1, Issue 3, Spring/Summer 2012



INTERNATIONAL NEURAL NETWORK SOCIETY

Retail: US$10.00

Natural Intelligence: the INNS Magazine
Editorial Board

The International Neural Networks Society (INNS) is embarking on a new journey. Not satisfied with its own past successes, INNS is constantly looking for new ways to better itself. The goal is for INNS to be the most prestigious professional organization in fields around neural networks and natural intelligence (broadly defined), as it has been for years. To keep up with the fast changing world of relevant science and technology, a new magazine that is designed to appeal to a broader readership ---the new INNS magazine entitled "Natural Intelligence"---thus is born.

Ron Sun, President of the International Neural Networks Society

The new INNS magazine aims at bridging different communities, spreading from neuroscientists to information engineers, and also from university students to world leading researchers. We define "Natural Intelligence" to include both "intelligence existing in nature" and "intelligence based on the state of things in nature". Therefore, the new INNS magazine "Natural Intelligence" plans to cover (a) experiments, (b) computational models, and (c) applications of the intelligent functions in our brains. Also, there is an important need for well-written introductory papers targeting both young and established researchers from other academic backgrounds. The interdisciplinary nature of the many new emerging topics makes these introductory papers essential for research on Natural Intelligence. Therefore, the new INNS magazine will mainly publish (a) review papers, (b) white papers, and (c) tutorials. In addition, columns, news, and reports on the communities will also be included.

Soo-Young Lee, Editor-in-Chief, Natural Intelligence: the IN INNS Magazine

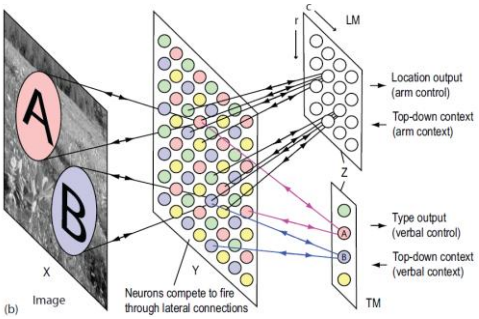
Spring and Summer issues are combined as the 3rd issue of Volume 1, and the 4th issue will be online at Fall 2012. Then, the second volume is scheduled from January 2013.

Natural Intelligence

ISSN 2164-8522

The INNS Magazine

Volume 1, Issue 3, Spring/Summer 2012



Natural Intelligence: the INNS Magazine is published quarterly by the International Neural Network Society (INNS) at *www.inns.org* and *www.ni-inns.info*. **Headquarters**: 2424 American Lane, Madison, WI 53704, U.S.A. Telephone: +1-608-443-2461. Fax: +1-608-443-2474 or 1-608-443-2478. E-mail addresses: inns@reesgroupinc.com.
All submission should be made to *inns.ni@gmail.com* or *ni@neuron.kaist.ac.kr*.

Understanding Human Implicit Intention

Soo-Young Lee
Editor-in-Chief, Natural Intelligence: the INNS Magazine

The understanding of human implicit intention is an interesting topic in cognitive neuroscience, and its applications may open a new horizon for the intelligent human-machine interface. The current user interface has been developed to understand the explicit representation of human intention such as keystrokes, gestures, and speech for appropriate responses. However, there exist many cases in which people do not show their intention explicitly. Even the actual intention may be different from the explicit one. Therefore, for the next-generation intelligent human-computer interface, it becomes very important to understand the 'implicit' intention which includes both the 'intentionally-hidden' intention and 'un-represented' intention. Although the former has been investigated in connection with the lie detection, the latter is yet to be investigated.

Recently several researches were reported on the understanding of un-represented intentions. However, these have been limited to specific applications such as web surfing and motion-based intentions. More genetic definition of implicit intention components is necessary. For example, the edges and the frequency are the basic components of vision and auditory perception, respectively. Also, people in general agree with the basic components of human emotions, i.e., happiness, sadness, disgusting, etc. We propose to define 'sympathy for the other' and 'non-sympathy for the other' as two basic components of the un-represented implicit intentions. Since the machine needs to understand human intention during human-machine interaction, the above definition is quite meaningful.

The main difficulty of the researches on the implicit intention resides in the non-existence of the ground truth. Therefore, for relatively obvious experimental conditions, it may be advantageous to measure multimodal signals such as fMRI, EEG, eye-tracking, pupil dilation, GSR, audio and visual signals. A binary classifier may be trained. Then, the trained classifier may be used to understand implicit intentions for less obvious conditions in real-world applications.

Soon machine will understand human intentions, both explicit and implicit, and provide appropriate services for human. You do not need show your intention. Also, you cannot hide your intention. However, we will have a 'good' big brother to serve us.

Intelligent to machine, freedom to mankind! ■

A Computational Introduction to the Biological Brain-Mind

Juyang Weng
Michigan State University, USA
**corresponding author*: weng@cse.msu.edu

Abstract

The brain-mind is hyphened because of the tight integration of the brain model and the mind model. On one hand, the neuroscience literature has provided very rich data about the brain, but such data tend to mislead us to think that the brain is composed many special purpose modules (e.g., Brodmann areas) where the role of each module is largely determined by the genes (e.g., detecting edge orientation or a human face). On the other hand, traditional artificial neural networks (e.g., Self-Organization Map SOM) perform general-purpose signal processing and they learn. However, they cannot autonomously learn and develop like a brain with its body. Autonomous mental development models how a brain-like system, natural and artificial, develops autonomously through interactions with the environments. The most fundamental difference between traditional machine leaning (using symbolic or neural net methods) and autonomous mental development is that a developmental program is task non-specific so that it can autonomously generate internal representations for a wide variety of simple to complex tasks. This paper first discusses why autonomous development is necessary based on a concept called task muddiness. No traditional methods can perform muddy tasks. If the electronic system that you design is meant to perform a muddy task, you need to enable it to autonomously develop its own mind. Then some basic concepts of autonomous development are explained, including the paradigm for autonomous development, brain-mental architectures, developmental algorithm, a refined classification of types of machine learning, spatial complexity and time complexity. Finally, the architecture of a brain-like spatiotemporal machine that is capable of autonomous development is described.

1. Biological Development

A human being starts to develop from the time of conception. At that time, a single cell called a zygote is formed. In biology, the term *genotype* refers to all or part of the genetic constitution of an organism. The term *phenotype* refers to all or part of the visible properties of an organism that are produced through the interaction between the genotype and the environment. In the zygote, all the genetic constitution is called genome, which mostly resides in the nucleus of a cell. At the conception of a new human life, a biological program called the *developmental program* starts to run. The code of this program is the genome, but this program needs the entire cell as well as the cell's environment to run properly.

The biological developmental program handles two types of development, *body development* and *mental development*. The former is the development of everything in the body excluding the brain. The latter is the development of the brain (or the Central Nervous System CNS). Through the body development, a normal child grows in size and weight, along with many other physical changes. Through the mental development, a normal child develops a series of mental capabilities through interactions with the environment. Mental capabilities refer to all known brain capabilities, which include, but not limited to, perceptual, cognitive, behavioral and motivational capabilities. In this paper, the term *development* refers to mental development unless stated otherwise. The biological mental development takes place in concurrence with the body development and they are closely related. For example, if the eyes are not normally developed, the development of the visual capabilities is greatly affected. In the development of an artificial agent, the body can be designed and fixed (not autonomously developed), which helps to reduce the complexity of the autonomous mental development.

The *genomic equivalence* principle [1] is a very important biological concept for us to understand how biological development is regulated. This principle states that the set of genes in the nucleus of every cell (not only that in the zygote!) is functionally complete -- sufficient to regulate the development from a single cell into an entire adult life. This principle is dramatically demonstrated by cloning. This means that there are no genes that are devoted to more than one cell as a whole. Therefore, development guided by the genome is *cell-centered*. Carrying a complete set of genes and acting as an autonomous machine, every cell must handle its own learning while interacting with its external environment (e.g., other cells). Inside the brain, every neuron develops and learns in place. It does not need any dedicated learner outside the neuron. For example, it does not need an extra-cellular learner to compute the covariance matrix (or any other moment matrix or partial derivatives) of its input lines and store extra-cellularly. If an artificial developmental program develops every artificial neuron based on only information that is available to the neuron itself (e.g., the cellular environment such as pre-synaptic activities, the developmental program inside the cell, and other information that can be biologically stored intra-cellularly), we call this type of learning *in-place learning*.

This in-place concept is more restrictive than a common concept called "local learning." For example, a local learning algorithm may require the computation of the covariance matrix of the pre-synaptic vector that must store extra-cellularly. In electronics, the in-place learning principle can greatly reduce the required electronics and storage space, in addition to the biological plausibility. For example, suppose that every biological neurons requires the partial derivative matrix of its pre-synaptic vector. As the average number of synapses of a neuron in the brain is on the order of $n = 1000$. Each neuron requires about $n^2 = 1,000,000$ storage units outside every neuron. This corresponds to about 1,000,000 of the total number of synapses (10^{14}) in the brain!

Conceptually, the fate and function of a neuron is not determined by a "hand-designed" (i.e., genome specified meaning of the external environment. This is another consequence of the genomic equivalence principle. The genome in each cell regulates the cells mitosis, differentiation, migration, branching, and connections, but it does not regulate the meaning of what the cell does when it receives signals from other connected cells. For example, we can fin a V1 cell (neuron) that responds to an edge of a particular orientation. This is just a facet of many emergent properties of the cell that are consequences of the cells own biological properties and the activities of its environment. A developmental program does not need to, and should not, specify which neuron detects a pre-specifie feature type (such as an edge or motion).

2. Why Autonomous Mental Development?

One can see that biological development is very "low level", regulating only individual neurons. Then, why is it necessary to enable our complex electronic machines to develop autonomously? Why do we not design high-level concepts into the machines and enable them to carry out our high-level directives? In fact, this is exactly many symbolic methods have been doing for many years. Unfortunately, the resulting machines are brittle — they fail miserably in real world when the environment fall out of the domains that have been modeled by the programmer.

To appreciate what are faced by a machine to carry out a complex task, Weng [2] introduced a concept called *task muddiness*. The composite muddiness of a task is a multiplicative product of many individual muddiness measures. There are many possible individual muddiness measures. Those individual muddiness measures are not necessarily mutually independent or at the same level of abstraction, since such a requirement is not practical nor necessary for describing the muddiness of a task. They fall into five categories: (1) external environment, (2) input, (3) internal environment, (4) output and (5) goal, as shown in Table I. The term "external" means external with respect to the brain and "internal" means internal to the brain.

TABLE I

A LIST OF MUDDINESS FACTORS FOR A TASK

Category	Factor	Clean ⟷	Muddy
External Env.	Awareness	Known	Unknown
	Complexity	Simple	Complex
	Controlledness	Controlled	Uncontrolled
	Variation	Fixed	Changing
	Foreseeability	Foreseeable	Nonforeseeable
Input	Rawness	Symbolic	Real sensor
	Size	Small	Large
	Background	None	Complex
	Variation	Simple	Complex
	Occlusion	None	Severe
	Activeness	Passive	Active
	Modality	Simple	Complex
	Multi-modality	Single	Multiple
Internal Env.	Size	Small	Large
	Representation	Given	Not given
	Observability	Observable	Unobservable
	Imposability	Imposable	Nonimposable
	Time coverage	Simple	Complex
Output	Terminalness	Low	High
	Size	Small	Large
	Modality	Simple	Complex
	Multimodality	Single	Multiple
Goal	Richness	Low	High
	Variability	Fixed	Variable
	Availability	Given	Unknown
	Telling-mode	Text	Multimodal
	Conveying-mode	Simple	Complex

The composite muddiness of a task can be considered as a product of all individual muddiness measures. In other words, a task is extremely muddy when all the five categories have a high measure. A chess playing task with symbolic input and output is a clean problem because it is low in categories (1) through (5). A symbolic language translation problem is low in (1), (2) and (4), moderate in (3) but high in (5). A vision-guided navigation task for natural human environment is high in (1), (2), (3) and (5), but moderate in (4). A human adult handles extremely muddy tasks that are high in all the five categories.

From the muddiness table Table I we have a more detailed appreciation what a human adult deals with even in a daily task, e.g., navigating or driving in a city environment. The composite muddiness of many tasks that a human or a machine can execute is proposed by Weng [2] as a metric for measuring required intelligence.

A human infant is not able to perform those muddy tasks that a human adult performs everyday. The process of mental development is necessary to develop such a wide array of mental skills. Much evidence in developmental psychology has demonstrated that not only a process of development is necessary for human intelligence, the environment of the development is also critical for normal development.

Likewise, it is not practical for a human programmer to program a machine to successfully execute a muddy task. Computers have done very well for clean tasks, such as playing chess games. But they have done poorly in performing muddy tasks, such as visual and language understanding. Enabling a machine to autonomously develop task skills in

its real task environments is the only approach that has been proved successful for muddy tasks — no existing higher intelligence for muddy tasks is not developed autonomously.

3. The Paradigm of Autonomous Development

By definition an agent is something that senses and acts. A robot is an agent, so is a human. In the early days of artificia intelligence, smart systems that caught the general public's imagination were programmed by a set of task-specifi rules. The file of artificia intelligence moved beyond that early stage when it started the trend of studying general agent methodology [3], although the agent is still a task-specifi machine.

As far as we know, Cresceptron 1993 [4], [5] was the firs developmental model for visual learning from complex natural backgrounds. By developmental, we mean that the internal representation fully emerges from interactions with environment, without allowing a human to manually instantiate a task-specifi representation. By the mid 1990's, connectionists had started the exploration of the challenging domain of development [6], [7], [8].

Due to a lack of the breadth and depth of the multi-disciplinary knowledge in the single mind of a researcher or a reviewer, there have been various doubts from domain experts, mainly due to the widespread lack of sufficien cross-disciplinary knowledge discussed above. Examples include: (1) Artificia intelligence does not need to follow the brain's way. (2) Modeling the human mind does not need to follow the brain's way. (3) Your commitment to understanding the brain is laudable but naive; (4) Regardless a clear advance of knowledge and its importance, I need to see X.

There is a lack of bylaws, guidelines and due process that contain the negative effects of human nature that are well documented by Thomas Kuhn [9]. Such negative effects eroded "revolutionary advances" required by some federal programs. Serious overhauls and investments for the infrastructure for converging research on intelligence is urgently needed. Such infrastructure is necessary for the healthy development of science and technology in the modern time.

Not until the birth of the new AMD fiel marked by the NSF and DARPA funded Workshop on Development and Learning 2000 [10], [11] has the concept of the task-nonspecifi developmental program caught the attention of many researchers. A hallmark difference between traditional artificia intelligence approaches and autonomous mental development [11] is the task specificit . All the existing approaches to artificia intelligence is task specific except the developmental approach. Table II lists the major differences among existing approaches to artificia intelligence. An entry marked as "avoid modeling" means that the representation is emergent from experience. See Weng 2012 [12] for a review of symbolic models and emergent models and the comparison thereof.

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-specifi representation. Depending on different approaches, different techniques are used to produce the mapping from sensory inputs to effector outputs. The techniques used range from direct programming (knowledge-based approach), to learning the parameters (in the parametric model), to genetic search (genetic approach). Although genetic search is a powerful method, the chromosome representations used in artificia genetic search algorithms are task specific

Using the developmental approach, 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 refl xes 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 practical issue.

4. Learning Types

In the machine learning literature, there have been widely accepted definition of learning types, such as supervised, unsupervised, and reinforcement learning. However, these conventional definition are too coarse to describe computational learning through autonomous development. For example, it is difficul to identify any type of learning that is completely unsupervised. Further, the traditional classificatio of animal learning models, such as classical conditioning and instrumental conditioning, is not sufficien to address computational considerations of every time instant of learning. A definitio of a refine classificatio of learning types is necessary.

We use a variable i to indicate *i*nternal task-specifi representation imposed by human programmer (called *internal-state imposed* $i = 1$) or not (called *internal-state autonomous* $i = 0$).

We use e to denote autonomy of effector. If the concerned effector is directly guided by the human teacher or other teaching mechanisms for the desired action, we call the situation *action imposed* ($e = 1$). Otherwise, the learning is effector autonomous ($e = 0$).

TABLE II

A COMPARISON OF APPROACHES TO ARTIFICIAL INTELLIGENCE

Approach	Species Architecture	World knowledge	Agent behaviors	Task specifi
Knowledge-based	Model	Model	Model	Yes
Learning-based	Model	Parametrically model	Model	Yes
Behavior-based	Model	Avoid modeling	Model	Yes
Genetic	Genetic search	Parametrically model	Model	Yes
Developmental	Parametrically model	Avoid modeling	Minimize modeling	No

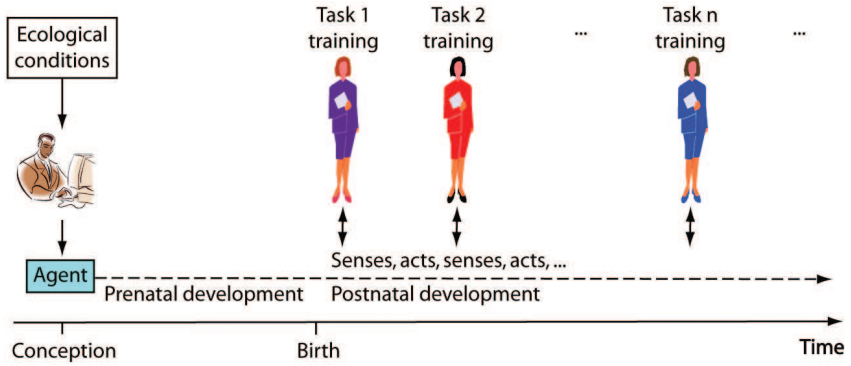


Figure 1. Illustration of the paradigm of developmental agents, inspired by human mental development. No task is given during the programming (i.e., conception) time, during which a general-purpose task-nonspecifi developmental program is loaded onto the agent. Prenatal development is used for developing 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.

We need to distinguish the channels of reward (e.g., sweet and pain sensors) that is available at the birth time, and other channels of reward that are not ready to be used as reward at the birth time (e.g., auditory input "good" or "bad") but implies a value after a certain amount of development. We defin (inborn) biased sensors:

If the machine has a predefine preference pattern to the signals from a sensor at the birth time, this sensor is an *(inborn) biased sensor*. Otherwise, it is an *(inborn) unbiased sensor*.

In fact, all the sensors become biased gradually through postnatal experience — the development of the value system. For example, the image of a fl wer does not give a newborn baby much reward, but the same image becomes pleasant to look at (high value) after the baby has gown up.

We use the third variable b to denote whether a biased sensor is used. If any biased sensor is activated (sensed) during the learning, we called the situation *reinforcement* ($b = 1$). Otherwise, the learning is called *communicative* ($b = 0$).

Using these three key factors, any type of learning can be represented by a 3-tuple (i, e, b), which contains three components i, e, and b, each of which can be either represented by 0 or 1. Thus, there are a total of 8 different 3-tuples, representing a total of 8 different learning types. If we consider ieb as three binary bits of the type index number of learning type, we have 8 types of learning define in Table III. We can also name each type. For example, Type 0 is state-autonomous, effector-autonomous, communicative learning. Type 7 is state-imposable, effector-imposed, reinforcement learning, but it has not been included in the traditional definitio of either supervised learning or reinforcement learning. However, this learning is useful when teaching a positive or negative lesson through supervision.

TABLE III

EIGHT TYPES OF BIOLOGICAL AND ARTIFICIAL LEARNING

Type (binary)	Internal state	Effector	Biased sensor
0 (000)	Autonomous	Autonomous	Communicative
1 (001)	Autonomous	Autonomous	Reinforcement
2 (010)	Autonomous	Imposed	Communicative
3 (011)	Autonomous	Imposed	Reinforcement
4 (100)	Imposable	Autonomous	Communicative
5 (101)	Imposable	Autonomous	Reinforcement
6 (110)	Imposable	Imposed	Communicative
7 (111)	Imposable	Imposed	Reinforcement

Using three key features, state-imposed, effector-imposed and reinforcement, eight learning types are defined This refine definitio is necessary to understanding various modes of developmental and nondevelopmental learning.

All learning types using a non-developmental learning method corresponding to Types 7 to 4. This is because the task-specifi representation is at least partially handcrafted after the task is given. Autonomous mental development uses Types 0 to 3.

5. Brain-Mind Architectures

Weng 2007 [13] proposed a SASE model through which the agent can autonomously learn to think, while the thinking behavior is manifested as internal attention. Attention is a key to emergent intelligence.

5.1. Top-down Attention is Hard

Consider a car in a complex urban street environment. Attention and recognition is a pair of dual-feedback problems. Without attention, recognition cannot do well; recognition requires attended areas (e.g., the car area) for the further processing (e.g., to recognize the car). Without recognition, attention cannot do well; attention requires recognition for guidance of the next fixatio (e.g., a possible car area).

1) Bottom-up attention: Studies in psychology, physiology, and neuroscience provided qualitative models for bottom-up attention, i.e., attention uses different properties of sensory inputs, e.g., color, shape, and illuminance to extract saliency. Several models of bottom-up attention have been published. The firs explicit computational model of bottom-up attention was proposed by Koch & Ullman in 1985 [14], in which a "saliency map" is computed to encode stimuli saliency at every lactation in the visual scene. More recently, Itti & Koch et al. 1998 [15] integrated color, intensity, and orientation as basic features in multiple scales for attention control. An active-vision system, called NAVIS (Neural Active Vision) by Baker et al. 2001, was proposed to conduct the visual attention selection in a dynamic visual scene [16]. Our SASE model to be discussed next indicates that saliency is not necessarily independent of learning: The top-down process in the previous time instant may affect the current bottom-up saliency.

2) Top-down attention: Volitional shifts of attention are also thought to be performed top-down, through spatially define and feature-dependant controls. Olshausen et al. 1993 [17] proposed a model of how visual attention can be directed to address the position and scale invariance in object recognition, assuming that the position and size information is available from the top control. Tsotsos et al. 1995 [18] implemented a version of attention selection using a combination of a bottom-up feature extraction scheme and a top-down position selective tuning scheme. Rao et al. 2004 [19] described a pair of cooperating neural networks, to estimate object identity and object transformations, respectively. Schill et al. 2001 [20] presented a top-down, knowledge-based reasoning system with a low-level pre-processing where eye movement is to maximize the information about the scene. Deco & Rolls 2004 [21] wrote a model of object recognition that incorporates top-down attention mechanisms on a hierarchically organized set of visual cortical areas. In the above studies, the model of Deco & Rolls 2004 [21] was probably the most biologically plausible, as it incorporates bottom-up and top-down fl ws into individual neuronal computation, but unfortunately the top-down connections were disabled during learning and no recognition performance data were reported.

Where-What Networks (WWN are embodiment of a brain-mind model called Developmental Network (DN). In the Where-What Network 2 (WWN-2) experiment [22] discussed later, we found that the corresponding network that drops the L4-L2/3 laminar structure gave a recognition rate lower than 50%. In other words, a network that treats top-down connection similar to bottom-up connection (like a uniform liquid state machine [23]) is not likely to achieve an acceptable performance.

5.2. Motor Shapes Cortical Areas

On one hand, high-order (i.e., later) visual cortex of the adult brain includes functionally specifi regions that preferentially respond to objects, faces, or places. For example, the fusiform face area (FFA) responds to face stimuli (Kanawisher 1997 [24], 1999 [25], Grill-Spector et al. 2004 [26]) and the parahippocampal place area (PPA) responds to place identity (O'Keefe & Dostrovsky 1971 [27], Ekstrom et al. 2003 [28], Bohbot & Corkin 2007 [29]). How does the brain accomplish this feat of localizing internal representation based on meaning? Why is such a representation necessary?

In the cerebral cortex, there is a dense web of anatomically prominent feedback (i.e., top-down) connections (Kennedy & Bullier 1985 [30], Perkel et al. 1986 [31], Felleman & Van Essen 1991, [32], Katz & Callaway 1992 [33], Salin & Bullier 1995 [34], Johnson & Burkhalter 1996 [35]). It has been reported that cortical feedback improves discrimination between figur and background and plays a role in attention and memory (Hupe et al. 1998 [36], Grossberg & Raizada 2000 [37], Sullivan & de Sa [38]). Do feedback connections perform attention? Furthermore, do feedback connections play a role in developing abstractive internal representation?

The computational roles of feedback connections in developing meaning-based internal representations have not been clarifie in existing studies reviewed above. The Self-Abstractive Architecture next indicates that in the cerebral cortex, each function layer (L4 and L2/3) is a state at this layer. We will show that, unlike the states in POMDP, HMM, Hopfiel network and many others, the states in the Self-Abstractive Architecture integrate information from bottom-up inputs (feature inputs), lateral inputs (collaborative context) and top-down inputs (abstract contexts) into a concise continuous vector representation, without the artificia boundaries of a symbolic representation.

5.3. Brain Scale: "Where" and "What" Pathways

Since the work of Ungerleider and Mishkin 1982 [39], [40], a widely accepted description of visual cortical areas is illustrated in Fig. 2 [32], [17]. A ventral or "what" stream that runs from V1, to V2, V4, and IT areas TEO and TE

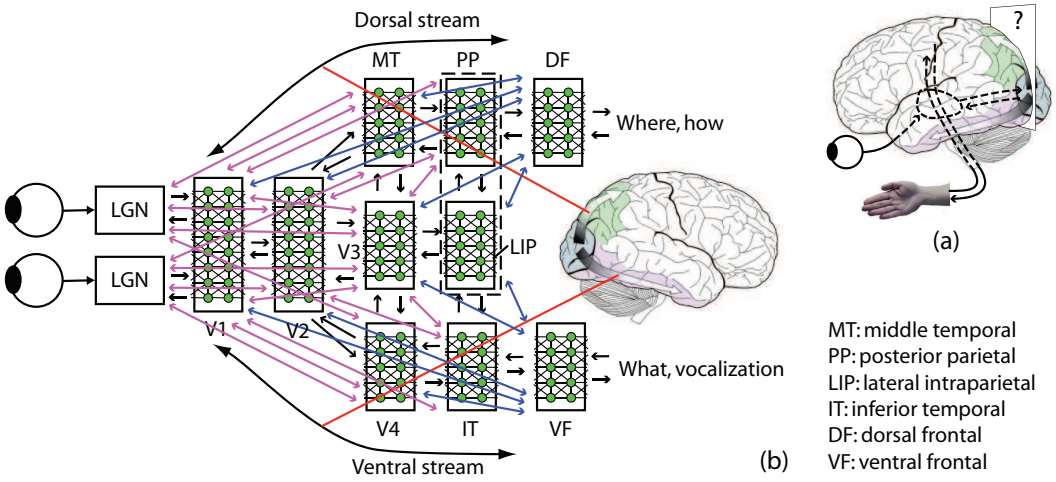


Figure 2. (a) How does the brain generate internal representation? The imaginary page slices the brain to "peek" into its internal representation. The only external sources are sensors and effectors. They are the ports for the brain to exchange information with the external environment. (b) An example of brain connections — the visuomotor streams. It consists of two major streams among others: the dorsal "where and how" stream and the ventral "what" stream. The nature of the processing along each stream is shaped by not only sensory input-output but also the motor input-output. In principle, every area needs connections with all other areas. An area pair that has only weak connection means that this pair has only weak statistical correlation. This diagram only schematically illustrates the cortical connection patterns. Every two-way arrow means two one-way arrows in opposite directions.

computes properties of object identity such as shape and color. A dorsal or "where" stream that runs from V1, to V2, V3, MT and the medial superior temporal areas MST, and on to the posterior parietal cortex (PP) computes properties of the location of the stimulus on the retina or with respect to the animal's head. Neurons in early visual areas have small spatial receptive field (RFs) and code basic image features; neurons in later areas have large RFs and code abstract features such as behavioral relevance. Selective attention coordinates the activity of neurons to affect their competition and link distributed object representations to behaviors (e.g., see the review by Serences and Yantius 2006 [41]).

With the above rich, suggestive information from neuroscience, I propose that the development of the functions of the "where" and "what" pathways is largely due to:

1) Downstream motors. The motor ends of the dorsal pathway that perform position tasks (e.g., stretching an arm to reaching for an apple or a tool), and the motor ends of the ventral pathway that perform type classificatio and conceptual tasks (e.g., different limbic needs between a food and an enemy);
2) Top-down connections. The top-down connections from motor areas that shape the corresponding pathway representations.

Put in a short way, *motor is often abstract*. Any meaning that can be communicated between humans is motorized: spoken, written, hand-signed, etc. Of course, "motor is abstract" does not mean that every stage of every motor action sequence is abstract. However, the sequences of motor actions provide statistically crucial information for the development of internal abstractive representation.

5.4. System views

The system level architecture is illustrated in Fig. 3.

An agent, either biological or artificial can perform regression and classification

Regression: The agent takes a vector as input (a set of receptors). For vision, the input vector corresponds to a retinal image. The output of the network corresponds to motor signals, with multiple components to be active (firing) The brain is a very complex spatiotemporal regressor.

Classification: The agent can perform classificatio before it has developed sophisticated human language capability to verbally tell us the name of a class. For example, each neuron in the output layer corresponds to a different class.

1) Two signal sources: sensor and motor: The brain faces a major challenge as shown in Fig. 2(a). It does not have the luxury of having a human teacher to implant symbols into it, as the brain is not accessible directly to the external human teacher. Thus, it must generate internal representations from the two signal sources: the sensors and the effectors (motors). This challenging goal is accomplished by the brain's where-what networks schematically illustrated in Fig. 4. The system has two motor areas, the where motor that indicates where the attended object is and the what motor that tells what the attended object is. This specialization of each pathway makes computation of internal representation more effective.

5.5. Pathway Scale: Bottom-up and Top-down

It is known that cortical regions are typically interconnected in both directions [32], [42], [43]. However,

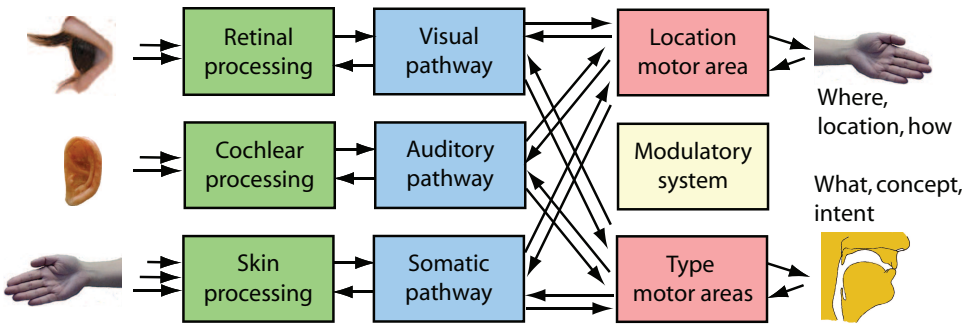


Figure 3. Schematic connections for multimodal integration. Multi-sensory and multi-effector integration is achieved through interactive learning. The modulatory (motivational) system is distributed over the entire brain through different types of neural transmitters. The block in the diagram only indicates the existence of the distributed modulatory system.

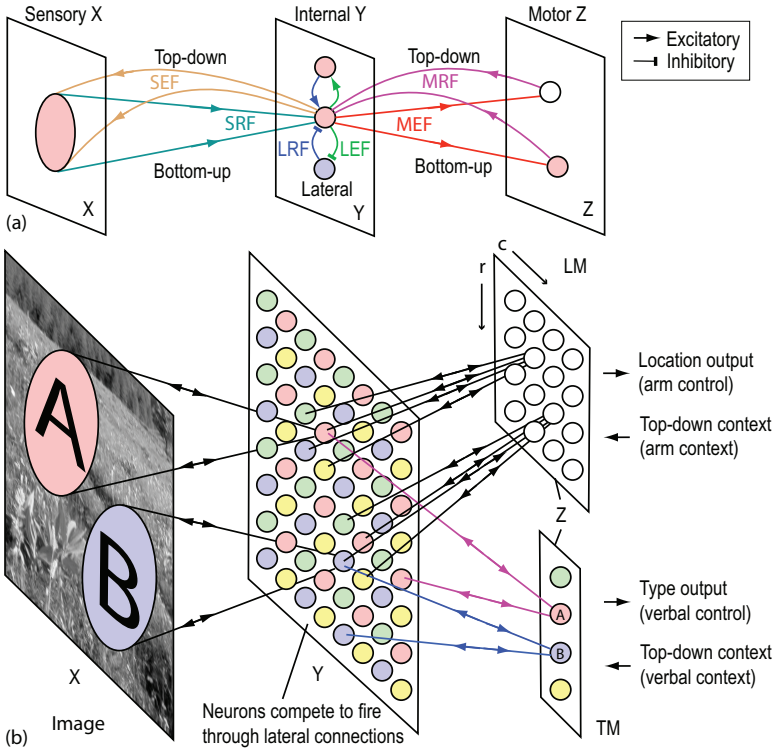


Figure 4. A simple WWN as a schematic developmental model of the brain. (a) The hextuple field for each neuron: SRF, MRF, LRF, SEF, MEF, and LEF, thus are highly recurrent. (b) A simple WWN with four areas (image as the X area, the brain as Y, and LM and TM as the Z area) and its hextuple network representation. Each wire connects if the pre-synaptic and post-synaptic neurons have co-fired The weight is the frequency of pre-synaptic co-firin when the post-synaptic neuron fires Within each cortical area, each neuron connects with highly correlated neurons using excitatory connections (e.g., NMDA-ergic) but connect with highly anti-correlated neurons using inhibitory connections (e.g., GABA-ergic). This forces neurons in the same area to detect different features in SRF and MRF. These developmental mechanisms result in the shown connections. Every Y neuron is *location-specific* and *type-specific*, corresponding to an object type (marked by its color) and to a location block (2×2 size each). Each LM neuron is location-specifi and type-invariant (more invariance, e.g., lighting-direction invariance, in more mature WWNs). Each TM neuron is type-specifi and location-invariant (more invariance in more mature WWNs). Each motor neuron pulls all applicable cases from Y. It also top-down boosts all applicable cases in Y as top-down context. A two-way arrow means two one-way connections. With each area, all the connections within the same area are omitted for clarity.

computational models that incorporate both bottom-up and top-down connections have resisted full analysis [44], [45], [46], [21], [47], [48], [49]. The computational model, illustrated in Fig. 5, provides further details about how each functional level in cortex takes inputs from the bottom-up signal representation space X and top-down signal representation space Z to generate and update self-organized cortical **bridge representation** space Y. This model further

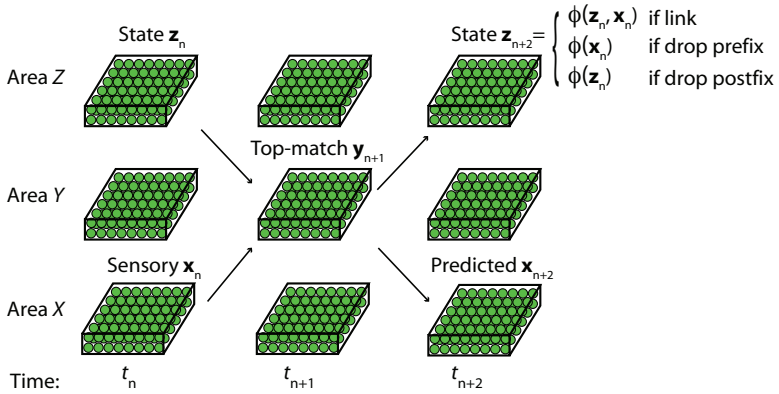


Figure 5. Cortex scale: The spatial SASE network for both spatial processing and temporal processing without dedicated temporal components. At each temporal unit shown above (two time frames), three basic operations are possible: link, drop prefix and drop postfix After proper training, the TCM is able to attend any possible temporal context up to the temporal sampling resolution.

computationally predicts that a primary reason for the dorsal and ventral pathways to be able to deal with "where" and "what" (or achieving identity and positional invariances [49]), respectively, is that they receive top-down signals that drive their motors.

From where does the forebrain receive teaching signals that supervise its motors? Such supervised-motor signals can be generated either externally (e.g., a child passively learns writing while his teacher manually guides his hand) or internally (e.g., from the trials generated by the spinal cord or the mid brain). As illustrated in Fig. 4, the model indicates that from early to later cortical areas, the neurons gradually increase their receptive field and gradually reduce their effective field as the processing of the corresponding *bridge representations* becomes less sensory and more motoric.

5.6. Cortex Scale: Feature Layers and Assistant Layers

The cerebral cortex contains six layers: layer L1 is the superficia layer and layer L6 is the deep layer. Weng et al. 2008 [50] reasoned that L4 and L2/3 are two feature detection layers as shown in Fig. 5 with L5 assisting L2/3 and L6 assisting L4, in the sense of enabling long range lateral inhibition. Such long range inhibitions encourage different neurons to detect different features. The model illustrated in Fig. 5 was informed by the work of Felleman & Van Essen [32], Callaway and coworkers [43], [42], and others (e.g., [37]). There are no top-down connections from L2/3 to L4, indicating that L4 uses unsupervised learning (U) while L2/3 uses supervised (S) learning. Weng et al. 2008 [50] reported that such a *paired* hierarchy USUS led to better recognition rates than the unpaired SSSS alternative.

5.7. Level Scale: the Dually Optimal CCI LCA

As shown in Fig. 5, given parallel input space consisting of the bottom-up space X and the top-down input space Z, represented as $X \times Z$, the major developmental goal of each cortical level (L4 or L2/3 as two representative levels of each area in Fig. 5) is to have different neurons in the level to detect different features, but nearby neurons should detect similar features.

Each feature level faces two pairs of conflictin criteria which are probably implicit during biological evolution: (1) The spatial pair: with its limited number of neurons, the level must learn the best internal representation from the environment while keeping a stable long-term memory. (2) The spatiotemporal pair: with its limited child time for learning, the level must not only learn the best representation but also learn quickly without forgetting important mental skills acquired long time ago. The sparse coding principle [51] is useful to address the firs pair: Allowing only a few neurons (best matched) to fir and update. Other neurons in the level are long-term memory because they are not affected. In other words, in each cortical region, only closely related mental skills are replaced each time. Therefore, the role of each neuron as working memory or long-term memory is dynamic, depending on the feature match (i.e., binding) with the input, as shown in Fig. 6. However, this rough idea is not sufficien for optimality.

The cortex inspired Candid Incremental Covariance-free (CCI) Lobe Component Analysis (LCA) [52], [53] has the desired dual optimality: spatial and spatiotemporal, as illustrated in Fig. 6. CCI LCA models optimal self-organization by a cortical level with a limited resource: c neurons. The cortical level takes two parallel input spaces: the bottom-up space X and top-down space Z denoted as $P = X \times Z$ as illustrated by Fig. 5. Each input vector is then denoted as $\mathbf{p} = (\mathbf{x}, \mathbf{z})$ where $\mathbf{x} \in X$ and $\mathbf{z} \in Z$. CCI LCA computes c feature vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_c$. Associated with these c feature vectors is a partition of the input space P into c disjoint regions $R_1, R_2, ..., R_c$, so that the input space P is the union of all these regions. For the optimal distribution of neuronal resource, we consider that each input vector $\mathbf{p}$ is represented by the winner feature $\mathbf{v}_j$ which has the highest response r_j:

$$j = \arg \max_{1 \le i \le c} r_i$$

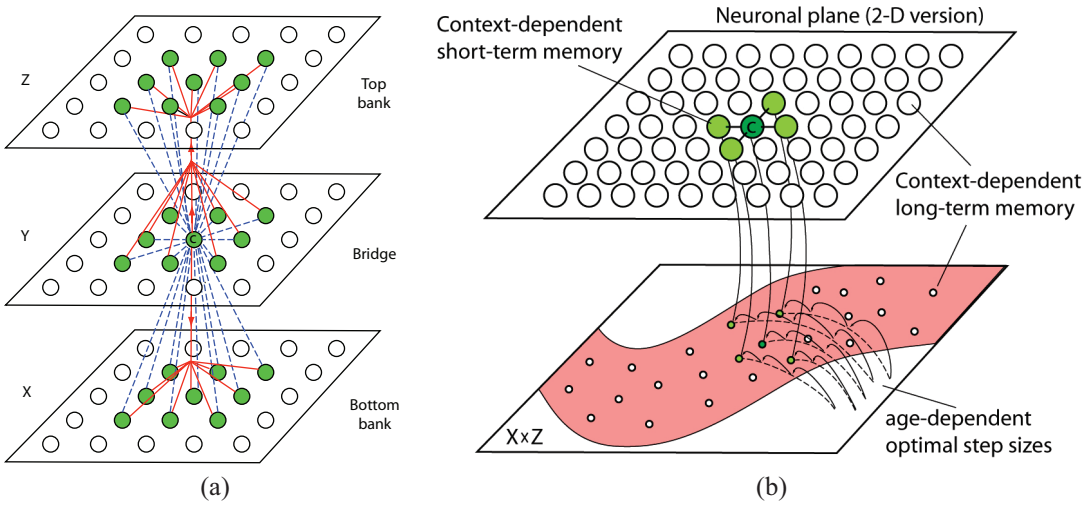


Figure 6. (a) The default connection pattern of every neuron in the brain. The area Y is a bridge for its two banks X and Z. The connections are local but two-way. Blue: neuronal input; red: axonal output. For each feature neuron (e.g., pyramidal neuron) in the brain, some near neurons (e.g., green for the center neuron) are connected to the neuron by excitatory connections (for prediction) and some far neurons (white ones) are connected to the center neuron by inhibitory connections (competition resulting in detection of different features by different neurons). Neurons that are not connected with the center neuron c are not considerably correlated or anti-correlated with it. (b) Cell-centered learning. The upper layer indicates the positions for the neurons in the same area: firin neurons are (context-dependent) working memory and those do not fir are (context dependent) long-term memory. The lower layer indicates the very high dimensional input space ($X \times Z$) of the area Y. The purple area in $X \times Z$ indicates the manifold of the input distribution. The connection curve from the upper neuron and lower small circle indicates the correspondence between the upper neuron and the feature that it detects. The neuronal weight vectors must quickly move to this manifold as the inputs are received and further the density of the neurons in the purple area should reflec the density of the input distribution. The challenge of learning and fast adaptation at various maturation stages of development is as follows: The updating trajectory of every neuron is a highly nonlinear trajectory. The statistical efficien y theory for neuronal weight update (amnesic average) results in the nearly minimum error in each age-dependent update, meaning not only the direction of each update is nearly optimal, but also every step length.

where r_i is the projection of input $\mathbf{p}$ onto the normalized feature vector $\mathbf{v}_i$: $r_i = \mathbf{p} \cdot (\mathbf{v}_i/\|\mathbf{v}_i\|)$. The form of approximation of $\mathbf{p}$ is represented by $\hat{\mathbf{p}} = r_i\mathbf{v}_i/\|\mathbf{v}_i\|$ and the error of this representation for $\mathbf{p}$ is $e(\mathbf{p}) = \|\hat{\mathbf{p}} - \mathbf{p}\|$.

1) Spatial optimality: The spatial optimality requires that the spatial resource distribution in the cortical level is optimal in minimizing the representational error. For this optimality, the cortical-level developmental program modeled by CCI LCA computes the best feature vectors $V = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_c)$ so that the expected square approximation error $\|\hat{\mathbf{p}}(V) - \mathbf{p}\|^2$ is statistically minimized:

$$V^* = (\mathbf{v}_1^*, \mathbf{v}_2^*, ..., \mathbf{v}_c^*) = \arg\min_V E\|\hat{\mathbf{p}}(V) - \mathbf{p}\|^2. \quad (1)$$

where E denotes statistical expectation. The minimum error means the optimal allocation of limited neuronal resource: frequent experience is assigned with more neurons (e.g., human face recognition) but rare experience is assigned with fewer neurons (e.g., fl wer recognition for a nonexpert). This optimization problem must be computed incrementally, because the brain receives sensorimotor experience incrementally. As the feature vectors are incrementally updated from experience, the winner neurons for the past inputs are not necessarily the same if past inputs are fed into the brain again (e.g., parents' speech when their baby was little is heard again by the grown-up baby). However, while the feature vectors are stabilized through extensive experience, the partition of the input space becomes also stable. Given a fi ed partition, it has been proved that the best feature set V^* consists of the c local firs principal component vectors, one for each region R_i. The term "local" means that the principal component vector for region R_i only considers the samples that fall into region R_i. As the partition is tracking a slowly changing environment (e.g., while the child grows up), the optimal feature set V^* tracks the slowly changing input distribution (called nonstationary random process).

Intuitively speaking, the spatial optimality means that with the same cortical size, all the children will eventually perform at the best level allowed by the cortical size. However, to reach the same mental skill level one child may require more teaching than another. The spatiotemporal optimality is deeper. It requires the best performance for every time t. That is, the child learns quickest allowed by the cortical size at every stage of his age.

2) Temporal optimality: The spatiotemporal optimality gives optimal step sizes of learning. Each neuron takes response weighted input $\mathbf{u}(t) = r(t)\mathbf{x}(t)$ at time t (i.e., Hebbian increment). From the mathematical theory of statistical efficien y, CCI LCA determines the optimal feature vectors $V^*(t) = (\mathbf{v}_1^*(t), \mathbf{v}_2^*(t), ..., \mathbf{v}_c^*(t))$ for every time instant t starting from the conception time $t = 0$, so that the distance

from $V^*(t)$ to its target V^* is minimized:

$$V^*(t) = \arg\min_{V(t)} E\|V(t) - V^*\|^2. \quad (2)$$

CCI LCA aims at this deeper optimality — the smallest average error from the starting time (birth of the network) up to the current time t, among all the possible estimators, under some regularity conditions. A closed form solution was found that automatically gives the optimal retention rate and the optimal learning rate (i.e., step size) at each synaptic update [52]

In summary, the spatial optimality leads to Hebbian incremental direction: response weighted pre-synamptic activity ($r\mathbf{p}$). The deeper spatiotemporal optimality leads to the best learning rates, automatically determined by the update age of each neuron. This is like different racers racing on a rough terrain along a self-determined trajectory toward an unknown target. The spatially optimal racers, guided by Hebbian directions, does not know step sizes. Thus, they cover other trajectories that require more steps. The spatiotemporally optimal racer, CCI LCA, correctly estimates not only the optimal direction at every step as illustrated in Fig. 6, but also the optimal step size at every step. In our experiments, CCI LCA out performed the Self-Organization Map (SOM) algorithm by an order (over 10 times) in terms of percentage distance covered from the initial estimate to the target. This work also predicts cell-age dependent plasticity schedule which needs to be verifie biologically.

6. Temporal Processing

Mauk & Buonomano 2004 [54] argued that the brain uses its intrinsic mechanisms to deal with time, and it does not have explicit delay lines and does not have a global clock. Drew & Abbott 2006 [55] proposed that the gradual change in the level of membrane potential inside a neuron may record some temporal information. However, this seems also not sufficien and robust for long time dependency, as argued by Ito et al. 2008 [56]. How the cortex deals with long time context has been elusive, especially considerably beyond around 30 ms modeled by Spike Timing-Dependent Plasticity (STDP) [57], [58].

To discuss how the brain deal with time, it is beneficia to discuss the Finite Automata (FA), also called finit state machines. In the symbolic world with vocabulary Σ, a sensory sequence of a life from conception up to the current time is a string $x \in \Sigma^*$, where Σ^* denotes the set of all strings of finit lengths. If the required skill set S for an agent can be define by a partition of Σ^* into c equivalent sets:

$$\Sigma^* = [q_1] \cup [q_2] \cup ... \cup [q_c]$$

where q_i is the equivalent perceptive-cognitive-behavior state, plus the state transitions among these states in the form of $q \xrightarrow{\sigma} q'$.

A Developmental Network (DN) is a generalization of Where-What Network whose motor area can represent any state: location, type, or any other cognitive state or behavior state or both. It has been established that given any FA, a DN can simulate any FA [59].

From the theory of FA [60], [61], the above conditions for state partition and state transitions are equivalent to the existence of the corresponding FA. It can be proved that a DN can simulate any FA. Further, during learning, the DN needs to learn every state transition of the FA only once.

Although the above discussion does not explicitly mention time, all the time properties are imbedded in the concept of equivalent states, such as the skills to estimate time duration, deal with time warping, conduct arbitrary temporal attention, and deal with context of any temporal length. Those properties are explicitly proved in Weng 2010 [62].

This completeness is symbolic since the environment of FA is symbolic. A major difference between a symbolic world and the real world is that the latter must deal with attention in new unobserved environments. Then, how a learned DN performs in a new environment and it can generalize depends on factors such as the similarity between a new environment and the learned environments. It is interesting here to utilize the power of the automata theory and the benefi of having mapped any arbitrary but static and symbolic FA to a general purpose but dynamic and emergent DN. Using the capability of DN to learn any Finite Automaton (FA), it can be proved that DN can abstract at least as well as FA-based symbolic models [63]. This addresses the correct criticisms by Marvin Minsky [64] and others in that traditional neural networks do not abstract well.

The DN has two types action and two types of sensing proposed by Weng 2007 [13]. Internal action is called *internal attention* here and external action is called *external behaviors*. Internal sensing is accomplished by autonomous internal wiring, *internal competition* and emergent representation. External sensing is realized by the sensors of the DN and its effectors (e.g., saccades, locomotion and limb actions which change what is sensed).

The external behaviors should include expressions about perception and cognition (e.g., speak), as well as actual actions acting on the external physical world. Although those skills are very different on surface, the DN model treats them in a unifie way so that a single unifie developmental program (DP) of a DN-based brain model potentially is sufficien to model complex skills. It is still unknown at this time what substantial limitations such a model has in modeling the human brain-mind.

7. Experiments

This introduction is theoretical. Experimental discussion is beyond the scope of this introduction. Some experimental results guided by this theory have published elsewhere. A visual WWN-2 [22] reached 92.5% in object recognition rate and 1.5 pixels in average position error with 75% of the area in each image fille with unknown natural

backgrounds. The WWN-3 [65] has shown a capability to deal with multiple learned objects in complex backgrounds. The user can specify either goal (location or type) and WWN-3 reports the reasoning results for other concepts (from location goal to type, or from type goal to location). WWN-4 [66] investigated internal connection constraints — it showed that deep learning as a cascade of areas between the sensory port and the motor port does not perform as well as shallow learning — multiple areas each having connections with both ports. WWH-5 [67] deals with objects with multiple scales in complex backgrounds.

WWN equipped with automatic synapse maintenance [68] demonstrated that how each neuron in WWN can segment object along its natural contour to get rid of the background in its sensory receptive field A temporal version for visual recognition [69] has reached an almost perfect recognition rate for centered objects viewed from any of the 360° object views. The stereo version of WWN [70] has shown that pre-screening is truly necessary for the temporal mechanisms to improve the result.

A text processing version [71] has been tested for part-of-speech tagging problem (assigning the words in a sentence to the corresponding part of speech, about 99% correct); and chunking (grouping sequences of words together and classify them by syntactic labels, about 96% success rate) using text corpus from the Wall Street Journal. A version of generative DN [72] has been shown to transfer skills from words through their temporal association, such as member-to-class transfer and member-to-member transfer.

8. Summary

The material in this paper outlines a series of tightly intertwined advances we recently made in understanding and modeling how the brain-mind develops and works. The grand picture of the biological brain-mind, although controversial and subject to further refinement seems getting increasingly clear. However, much work is needed to further verify whether such a grand model can generate rich behaviors that are consistent with known biological data and produce machine capabilities that traditional agent models cannot. A critical key that many computational models about cortical signal processing is *autonomous development*. Without understanding and modeling development, such models are not only incomplete, but also misleading, incorrect, and computationally inscalable. Developmental robots and machines urgently need industrial electronics for real-time, brain scale computation and learning. This need is here and now. This raises a great challenge for the fiel of industrial electronics, but an exciting future for understanding natural and artificia intelligence as well.

References

[1] W. K. Purves, D. Sadava, G. H. Orians, and H. C. Heller. *Life: The Science of Biology*. Sinauer, Sunderland, MA, 7 edition, 2004.

[2] J. Weng. Task muddiness, intelligence metrics, and the necessity of autonomous mental development. *Minds and Machines*, 19(1):93–115, 2009.

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

[4] 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'l Conf. Computer Vision*, pages 121–128, May 1993.

[5] J. Weng, N. Ahuja, and T. S. Huang. Learning recognition and segmentation using the Cresceptron. *International Journal of Computer Vision*, 25(2):109–143, Nov. 1997. Cited early versions in IJCNN 1992 and ICCV 1993.

[6] J. L. McClelland. The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. Bertelson, P. Eelen, and G. d'Ydewalle, editors, *International Perspectives on Psychological Science*, volume 1: Leading Themes, pages 57–88. Erlbaum, Hillsdale, New Jersey, 1994.

[7] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. *Rethinking Innateness: A connectionist perspective on development*. MIT Press, Cambridge, Massachusetts, 1997.

[8] S. Quartz and T. J. Sejnowski. The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20(4):537–596, 1997.

[9] T. S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, second edition, 1970.

[10] J. Weng and I. Stockman. Autonomous mental development: Workshop on development and learning. *AI Magazine*, 23(2):95–98, 2002.

[11] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.

[12] J. Weng. Symbolic models and emergent models: A review. *IEEE Trans. Autonomous Mental Development*, 4(1):29–53, 2012.

[13] J. Weng. On developmental mental architectures. *Neurocomputing*, 70(13-15):2303–2323, 2007.

[14] C. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.

[15] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.

[16] G. Backer, B. Mertsching, and M. Bollmann. Data- and model-driven gaze control for an active-vision system. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(12):1415–1429, December 2001.

[17] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11):4700–4719, 1993.

[18] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo Modeling visual attention via selective tuning. *Artificial Intelligence*, 78:507–545, 1995.

[19] R. P. N. Rao and D. H Ballard. Probabilistic models of attention based on iconic representations and predictive coding. In L. Itti, G. Rees, and J. Tsotsos, editors, *Neurobiology of Attention*. Academic Press, New York, 2004.

[20] K. Schill, E. Umkehrer, S. Beinlich, G. Krieger, and C. Zetzsche. Scene analysis with saccadic eye movemens: Top-down and bottom-up modeling. *Journal of Electronic Imaging*, 10(1):152–160, 2001.

[21] G. Deco and E. T. Rolls. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 40:2845–2859, 2004.

[22] Z. Ji and J. Weng. WWN-2: A biologically inspired neural network for concurrent visual attention and recognition. In *Proc. IEEE Int'l Joint Conference on Neural Networks*, pages +1–8, Barcelona, Spain, July 18-23 2010.

[23] M. Rabinovich, R. Huerta, and G. Laurent. Transient dynamics for neural processing. *Science*, 321:48–50, 2008.

[24] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11):4302–4311, 1997.

[25] N. Kanwisher, D. Stanley, and A. Harris. The fusiform face area is selective for faces not animals. *NeuroReport*, 10(1):183–187, 1999.

[26] K. Grill-Spector, N. Knouf, and N. Kanwisher. The fusiform face area subserves face perception, not generic within-category identification *Nature Neuroscience*, 7(5):555–562, 2004.

[27] J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, 1971.

[28] A. Ekstrom, M. Kahana, J. Caplan, T. Fields, E. Isham, E. Newman, and I. Fried. Cellular networks underlying human spatial navigation. *Nature*, 425:184–188., 2003.

[29] V. D. Bohbot and S. Corkin. Posterior parahippocampal place learning in h.m. *Hippocampus*, 17:863–872, 2007.

[30] H. Kennedy and J. Bullier. A double-labelling investigation of the afferent connectivity to cortical areas v1 and v2 of the macaque monkey. *Journal of Neuroscience*, 5(10):2815–2830, 1985.

[31] D. J. Perkel, J. Bullier, and H. Kennedy. Topography of the afferent connectivity of area 17 of the macaque monkey. *Journal of Computational Neuroscience*, 253(3):374–402, 1986.

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

[33] L. C. Katz and E. M. Callaway. Development of local circuits in mammalian visual cortex. *Annual Review of Neuroscience*, 15:31–56, 1992.

[34] P. A. Salin and J. Bullier. Corticocortical connections in the visual system: structure and function. *Physiological Review*, 75(1):107–154, 1995.

[35] R. R. Johnson and A. Burkhalter. Microcircuitry of forward and feedback connections within rat visual cortex. *Journal of Comparative neurology*, 368(3):383–398, 1996.

[36] J. M. Hupe, A. C. James, B. R. Payne, S. G. Lomber, P. Girard, and J. Bullier. Cortical feedback improves discrimination between figur and background by v1, v2 and v3 neurons. *Nature*, 394:784–787, Aug. 20 1998.

[37] S. Grossberg and R. Raizada. Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40:1413–1432, 2000.

[38] T. J. Sullivan and V. R. de Sa. A model of surround suppression through cortical feedback. *Neural Networks*, 19:564572, 2006.

[39] L. G. Ungerleider and M. Mishkin. Two cortical visual systems. In D. J. Ingel, editor, *Analysis of visual behavior*, pages 549–586. MIT Press, Cambridge, MA, 1982.

[40] M. Mishkin, L. G. Unterleider, and K. A. Macko. Object vision and space vision: Two cortical pathways. *Trends in Neuroscicence*, 6:414–417, 1983.

[41] J. T. Serences and S. Yantis. Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, 10(1):38–45, 2006.

[42] A. K. Wiser and E. M. Callaway. Contributions of individual layer 6 pyramidal neurons to local circuitry in macaque primary visual cortex. *Journal of neuroscience*, 16:2724–2739, 1996.

[43] E. M. Callaway. Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21:47–74, 1998.

[44] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastrate cortex. *Science*, 229(4715):782–784, 1985.

[45] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203, 2001.

[46] T. S. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, 20(7):1434–1448, 2003.

[47] M. D. Fox, M. Corbetta, A. Z. Snyder, J. L. Vincent, and M. E. Raichle. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. National Academy of Sciences U S A*, 103(26):10046–10051, 2006.

[48] T. J. Buschman and E. K. Miller. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315:1860–1862, 2007.

[49] 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'l Conference on Development and Learning*, pages 61–66, Monterey, CA, Aug. 9-12 2008.

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

[51] B. A. Olshaushen and D. J. Field. Emergence of simple-cell receptive fiel properties by learning a sparse code for natural images. *Nature*, 381:607–609, June 13 1996.

[52] J. Weng and N. Zhang. Optimal in-place learning and the lobe component analysis. In *Proc. IEEE World Congress on Computational Intelligence*, pages +1–8, Vancouver, BC, Canada, July 16-21 2006.

[53] J. Weng and M. Luciw. Dually optimal neuronal layers: Lobe component analysis. *IEEE Trans. Autonomous Mental Development*, 1(1):68–85, 2009.

[54] M. D. Mauk and D. V. Buonomano. The neural basis of temporal processing. *Annual Review of Neuroscience*, 27:307–340, 2004.

[55] P. J. Drew and L. F. Abbott. Extending the effects of spike-timing-dependent plasticity to behavioral timescales. *The National Academy of Sciences of the USA*, 103(23):8876–8881, 2006.

[56] I. Ito, R. C. Ong, B. Raman, and M. Stopfer. Sparse odor representation and olfactory learning. *Nature Neuroscience*, 11(10):1177–1184, 2008.

[57] G. Bi and M. Poo. Synaptic modificatio by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 24:139–166, 2001.

[58] Y. Dan and M. Poo. Spike timing-dependent plasticity: From synapses to perception. *Physiological Review*, 86:1033–1048, 2006.

[59] J. Weng. A 5-chunk developmental brain-mind network model for multiple events in complex backgrounds. In *Proc. Int'l Joint Conf. Neural Networks*, pages 1–8, Barcelona, Spain, July 18-23 2010.

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

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

[62] J. Weng. A general purpose brain model for developmental robots: The spatial brain for any temporal lengths. In *Proc. Workshop on Bio-Inspired Self-Organizing Robotic Systems, IEEE Int'l Conference on Robotics and Automation*, pages +1–6, Anchorage, Alaska, May 3-8 2010.

[63] J. Weng. Why have we passed "neural networks do not abstract well"? *Natural Intelligence: the INNS Magazine*, 1(1):13–22, 2011.

[64] M. Minsky. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Magazine*, 12(2):34–51, 1991.

[65] M. Luciw and J. Weng. Where What Network 3: Developmental top-down attention with multiple meaningful foregrounds. In *Proc. IEEE Int'l Joint Conference on Neural Networks*, pages 4233–4240, Barcelona, Spain, July 18-23 2010.

[66] M. Luciw and J. Weng. Where What Network 4: The effect of multiple internal areas. In *Proc. IEEE 9th Int'l Conference on Development and Learning*, pages 311–316, Ann Arbor,, August 18-21 2010.

[67] X. Song, W. Zhang, and J. Weng. Where-what network 5: Dealing with scales for objects in complex backgrounds. In *Proc. Int'l Joint Conference on Neural Networks*, pages 2795–2802, San Jose, CA, July 31 - August 5 2011.

[68] Y. Wang, X. Wu, and J. Weng. Synapse maintenance in the where-what network. In *Proc. Int'l Joint Conference on Neural Networks*, pages 2823–2829, San Jose, CA, July 31 - August 5 2011.

[69] M. Luciw, J. Weng, and S. Zeng. Motor initiated expectation through top-down connections as abstract context in a physical world. In *IEEE Int'l Conference on Development and Learning*, pages +1–6, Monterey, CA, Aug. 9-12 2008.

[70] M. Solgi and J. Weng. Developmental stereo: Emergence of disparity preference in models of visual cortex. *IEEE Trans. Autonomous Mental Development*, 1(4):238–252, 2009.

[71] J. Weng, Q. Zhang, M. Chi, and X. Xue. Complex text processing by the temporal context machines. In *Proc. IEEE 8th Int'l Conference on Development and Learning*, pages +1–8, Shanghai, China, June 4-7 2009.

[72] K. Miyan and J. Weng. WWN-Text: Cortex-like language acquisition with What and Where. In *Proc. IEEE 9th Int'l Conference on Development and Learning*, pages 280–285, Ann Arbor, August 18-21 2010.

Review Paper

Challenges for Brain Emulation: Why is it so Difficult?

Rick Cattell[1] and Alice Parker[2]

[1] SynapticLink.org
[2] University of Southern California, USA
**corresponding author:* rick@cattell.net

Abstract

In recent years, half a dozen major research groups have simulated or constructed sizeable networks of artificial neurons, with the ultimate goal to emulate the entire human brain. At this point, these projects are a long way from that goal: they typically simulate thousands of mammalian neurons, versus tens of billions in the human cortex, with less dense connectivity as well as less-complex neurons. While the outputs of the simulations demonstrate some features of biological neural networks, it is not clear how exact the artificial neurons and networks need to be to invoke system behavior identical to biological networks and it is not even clear how to prove that artificial neural network behavior is identical in any way to biological behavior. However, enough progress has been made to draw some conclusions and make comparisons between the leading projects. Some approaches are more scalable, some are more practical with current technologies, and some are more accurate in their emulation of biological neurons. In this paper, we examine the pros and cons of each approach and make some predictions about the future of artificial neural networks and the prospects for whole brain emulation.

Keywords: biomimetic, neuromorphic, electronic, artificial brain, neuron, intelligence

1. Introduction

Reverse-engineering the brain is one of the Grand Challenges posed by the United States National Academy of Engineering [1]. In this paper, we assess current status in approaching this difficult goal of brain emulation. We contrast competing approaches, and examine the major obstacles.

Artificial neurons and neural networks were proposed as far back as 1943, when Warren McColluch and Walter Pitts [2] proposed a "Threshold Logic Unit" with multiple weighted binary inputs combined to produce a binary output based on a threshold value. More sophisticated neural models were subsequently developed, including Rosenblatt's popular "perceptron" model [3] and others we examine in this article. In 1952, Hodgkin and Huxley [4] published a model of ionic currents that provided the first basis for mathematical modeling and simulation of biological neurons and their action potentials, with the help of Wilfred Rall's [5] theory of spatiotemporal integration, non-linear summation, and conductance of synaptic signals. These models have likewise been enhanced over the years by researchers examining synaptic transmission, integration, and plasticity.

Over the past 50 years, advances in technology have successively and phenomenally increased our ability to emulate neural networks with speed and accuracy.[1] At the same time, and particularly over the past 20 years, our understanding of neurons in the brain has increased substantially, with imaging and microprobes contributing significantly to our understanding of neural physiology.

These advances in both technology and neuroscience make possible the projects we discuss in this paper, aimed at modeling large numbers of interconnected neurons. Today it is feasible to emulate small but non-trivial portions of the brain, for example thousands of neurons in the visual cortex. Each approach has advantages and shortcomings when meeting the challenges posed by an artificial brain. We will examine the leading approaches and technologies, along with their pros and cons. We will conclude with a discussion of technological and architectural challenges for an artificial brain, and some debate on future research directions.

1.1. Motivation for Brain Emulation

Three motivations are frequently cited for brain emulation:

1. Researchers hope to gain a better understanding of how the brain works (and malfunctions) by creating simulations. A model can provide insight at all levels, from the biochemistry and neurochemical behavior of individual cells to the behavior of networks of neurons in the cortex and other parts of the brain.
2. Some researchers feel progress in artificial intelligence over the past 50 years has been insufficient to lead to intelligent behavior. Ideas from simulations of neural networks may yield new ideas to develop intelligent behavior in computers,

[1]Some authors refer to "simulating" neurons in software and "emulating" neurons in hardware, but for simplicity in this paper we use the term "emulation" to refer to hardware, software, and hybrid implementations.

for example through massive parallelism. Neural networks are already being used for applications such as computer vision and speech understanding, and many algorithmic approaches are bio-inspired, but their biological basis is, for the most part, simplified from the more-detailed models used by neuroscientists. Autonomous vehicles and other robotic applications are likely targets for such brain-like systems.

3. For the most part, computers still use the same basic architecture envisioned by John von Neumann in 1945. Hardware architectures based on the massive parallelism and adaptability of the brain may yield new computer architectures and micro-architectures that can be applied to problems currently intractable with conventional computing and networking architectures.

The projects described in this paper generally cite all three of these reasons for their work. However, there are differences in emphasis. Projects focused on understanding the brain require a more-detailed and more computationally-expensive model of neuron behavior, while projects aimed at the second or third goal may use simpler models of neurons and their connections that may not behave exactly as biological neural networks behave. An additional advantage of attempts at whole brain emulation is to further understanding of prosthetic device construction. While the research in that general area has focused on the difficult task of providing connectivity between electronics and biological neurons (e.g. Berger [6]), more complex emulated neural networks might one day provide prosthetic devices that adapt to an individual's brain, providing functions missing due to surgery, accidents or congenital defects.

1.2. Challenges to Brain Emulation

In spite of the progress in many brain emulation efforts, there are major challenges that must still be addressed:

- Neural complexity: In cortical neurons, synapses themselves vary widely, with ligand-gated and voltage-gated channels, receptive to a variety of transmitters [7]. Action potentials arriving at the synapses create post-synaptic potentials on the dendritic arbor that combine in a number of ways. Complex dendritic computations affect the probability and frequency of neural firing. These computations include linear, sublinear, and superlinear additions along with generation of dendritic spikes, and inhibitory computations that shunt internal cell voltage to resting potentials or decrease the potential, essentially subtracting voltage. Furthermore, some neuroscientists show evidence that the location of each synapse in the dendritic arbor is an important component of the dendritic computation [8], essential to their neural behavior, and there is growing consensus among neuroscientists that aspects of dendritic computation contribute significantly to cortical functioning. Further, some propagation of potentials and other signaling is in the reverse direction, affecting first-order neural behavior (for example, see the reset mechanism affecting dendritic spiking plasticity) [9, 10]. The extent of the detailed modeling of dendritic computations and spiking necessary for brain emulation is an open question.
- Scale: A massive system is required to emulate the brain: none of the projects we discuss have come close to this scale at present. The largest supercomputers and computer clusters today have thousands of processors, while the human cortex has tens of billions of neurons and a quadrillion synapses. We are a long way from cortex scale, even if one computer processor could emulate thousands of neurons, and, as we will see, it is unclear whether that emulation would be sufficiently accurate.
- Interconnectivity: Emulation of the cortex in hardware represents a massive "wiring" problem. Each synapse represents a distinct input to a neuron, and each postsynaptic neuron shares synapses with an average of 10,000 (and as many as 100,000) other presynaptic neurons. Similarly, the axon emerging from each neuronal cell body fans out to an average of 10,000 destinations. Thus each neuron has, on average, 10,000 inputs and 10,000 outputs. If the connections were mostly local, the wiring would not be so complicated; however, recent research by Bassett et al [11] derives a Rent exponent for the biological brain that could be used to compute the quantity of connections emerging from a volume of brain tissue. Early indications are that this Rent exponent is sufficiently large (many distal connections) so as to cause connectivity problems with conventional CMOS electronics.
- Plasticity: It is generally accepted that an emulated brain with static neural connections and neural behavior would not produce intelligence. Synapses must be "plastic": the strength of the excitatory or inhibitory connection must change with learning, and neurons must also be able to create new synapses and hence new connections during the learning process. Research on the mechanisms by which neurons learn, make and break connections, and possess memory is ongoing, with hypotheses and supporting data appearing frequently. These studies have led to a basic understanding of synaptic and structural plasticity. In the last decade, attention has been given to the role of glial cells in neural behavior, glial cells being much more numerous in the brain than neurons. The role of astrocytes, a type of glial cell, in learning and memory is being actively investigated[12] and neuromorphic circuits constructed [13].
- Power consumption: A final, indirect problem is the power consumed by a brain emulation with 50 billion neurons and 500 trillion connections, and the dissipation of the associated heat generated. The human brain evolved to use very little power, an estimated 25 watts. We do not have computing

technology anywhere near this power efficiency, although nanotechnology and ultra-low power design offer promise.

We will examine how each major project addresses these challenges. Although the brain emulation field is in its infancy, progress has been made in a very short time.

1.3 Other Surveys

Other surveys of brain emulation are worth reference here. They provide a different perspective than ours.

Sandberg and Bostrom [14] prove an excellent survey of the overall issues in brain emulation, although they do little discussion of actual brain emulation projects. They cover different levels of emulation, different neural models, computational requirements of emulation, and brain mapping technologies,

De Garis, Shuo, Goertzel, and Ruiting [15] provide the most similar survey to ours, covering half of the projects mentioned here. This is a good reference for another perspective on these projects. It is part one of a two-part survey. The second part, written by Goertzel, Lian, Arel, de Garis, and Chen [16], surveys higher-level brain models aimed at producing intelligent behavior, inspired by human intelligence but not based on emulation of neural networks; this work is closer to classical AI.

2. Background

There are three key components to any artificial neural network:

1. Neurons: the models used to emulate the computation and firing behavior of biological neurons, and the technology used for the emulation,
2. Connectivity: the models and technologies used for the synaptic connections between neurons, and
3. Plasticity: the models and technologies to create changes in the behavior of neurons and their synapses.

In this section, we provide some background on these models and technologies. This background will provide a basis for understanding brain emulation projects in the remainder of the paper.

2.1 Modeling Neurons

A variety of neural models are used in the projects we describe.

Most neural modeling involves the ion channels responsible for spike generation at the axon hillock, or the synapse, where spikes are transformed into post-synaptic potentials. The Hodgkin-Huxley [4] biological neural model discussed earlier, with Ca++, Na+, and K+ currents through ion channels, can require relatively expensive computations. Simulation is further complicated when one takes into account the 3-dimensional layout of axons and dendrites, requiring spatiotemporal integration. Cable theory and compartmental models have been used to account for the latter. Various improvements have been proposed to simplify computation while maintaining some level of faithfulness to biological neurons. A survey of this work is beyond the scope of this paper; the interested reader is referred to [17].

Some of the projects we discuss use a very simple model of neuronal behavior. The simplest model is an integrate-and-fire "point neuron," summing weighted input from synapses and comparing the resulting sum to a threshold, arriving at a binary decision whether and when to generate an output spike. This model is commonly extended to include a decaying charge, as a "leaky integrate and fire" neuron. The model can also be enhanced in other ways: non-linear summation, time-dependent thresholds, programmable delays in the delivery of spikes, and other variations. The point neuron models require only modest computation and hardware, in contrast to biological ion-channel models with spatiotemporal integration.

Izhikevich [18] provides a good recent survey of hybrid spiking neural models, comparing their computational costs and their ability to handle a range of 20 different spiking behaviors observed in neurons in vivo. Each model is represented by a set of ordinary differential equations that define the change in neuron membrane voltage over time, and the computational cost is measured by the number of floating point operations required in each time-step in the simulation. Izhikevich assumes 1 millisecond time steps in his paper. The Hodgkin-Huxley model is the most expensive one he considers, requiring about 1200 floating-point operations per millisecond.

Izhikevich concludes by advocating an enhanced leaky-integrate-and-fire model for neurons that is a compromise between computational cost and computational power, able to exhibit all 20 of the spiking waveforms he surveys. The differential equations for his model are

$$v' = .04\,v^2 + 5v + 140 - u + I, \quad (1)$$

$$u' = a\,(bv - u), \quad (2)$$

$$\text{if } v > 30 \text{ then reset } v \leftarrow c \text{ and } u \leftarrow u+d, \quad (3)$$

where a, b, c, d, and I are parameters that define the neuron's behavior, v is a variable representing the membrane potential in millivolts, and u is a variable representing membrane recovery. The parameter I represents the synaptic current resulting from the combination of post-synaptic potentials. Each millisecond of simulation requires only 13 floating-point operations in this model, about 100 times fewer floating point operations than Hodgkin-Huxley, yet the model still retains the capability to exhibit all of the same spiking behaviors as Hodgkin-Huxley, given appropriate values of the parameters.

More sophisticated neuron models, in contrast to the "point" models surveyed by Izhikevich, emulate components of the neuron separately. For example, synapses may be modeled separately from signal integration in the remainder of the neuron, followed by a spike-generator modeling the axon hillock, or a neuron may be modeled as dozens of small compartments, applying ion-migration equations to each compartment separately.

The fundamental feature of synapses is the voltage response over time of the neuron cell membrane to rapid input spikes that cause post-synaptic potentials to sum temporally and spatially, and that decay over time with time courses that vary depending on each individual synapse. The nonlinear sum of the excitatory post-synaptic potentials (EPSPs) might be offset by hyperpolarizing inhibitory post-synaptic potentials (IPSPs) that essentially subtract potential, or might be entirely negated by shunting inhibitory synapses that return the cell membrane to resting potential, with location of each synapse playing a role in the computation. The Blue Brain project we discuss models these dendritic computations in more detail than the other major projects.

As we shall see, the actual implementation of neuron models can be in software or in hardware, or a combination of the two. The purely-hardware implementations we discuss use neuromorphic analog circuits, as do the hardware portions of the hybrid implementations. We will discuss the pros and cons of these technology choices.

2.2 Modeling Connections

Modeling connections between neurons may seem trivial, given a hardware or software model of the neurons. However, one of the biggest challenges to brain emulation is the immense problem "wiring" the connections: the synapses, dendrites, and axons.

The connection-wiring problem differs depending how neurons are modeled and implemented. As we will see in the next section, three different approaches have been used to implement neurons:

1. *Supercomputers*, used to model neurons and their connections in software,
2. *Neuromorphic analog integrated circuits*, with an array of special-purpose neural-modeling circuits on each chip, and
3. *Special-purpose digital integrated circuits*, emulating neurons in software using many small CPUs networked together.

Corresponding to these neuron emulation technologies, there are several different approaches to implementing synaptic connectivity between neurons. In the supercomputer case, synaptic activity can be communicated through simple procedure calls or inter-process calls. In the case of neuromorphic analog circuits, direct wiring between artificial neurons has been used locally. However, since neurons contain many distinct synapses with differing effects on neural behavior, there is high connectivity fan-in for off-chip signals. As a result of the high connectivity fan-in and fan-out, with current technologies, direct wiring has only been practical for connections between "nearby" analog neurons. For longer connections in the analog case, and for all connections in the digital case, a networking approach has been required.

The approaches used for this networking in the major projects examined here are almost all based on Mahowald's pioneering *Address Event Representation* (AER) architecture[19]. Networking and AER are based on a simplifying assumption that continuous connectivity between neurons is not necessary for an accurate emulation. Instead, they assume communication is necessary only when a neuron fires, generating an action potential. The emulated neurons are networked together, generally with a topology of many nested networks, as on the Internet, to allow scaling. When a neuron fires, network packets are sent out to all of the neurons that synapse upon it, notifying them of the spike.

As on the Internet, each network node (a neuron in this case) is assigned a network-wide unique address, and some form of routing tables are required for the system to know what nodes and subnetworks a packet must go through to reach its destination. However, in typical network communication on the Internet, each network packet contains a source address, a destination address, and the data to be communicated. In contrast, the AER approach includes only the source address (the "address event" of the neuron that spiked) in the packet. A destination address is not used because it is not practical: every neuron would need to generate many thousands of packets each time it spiked.

Instead, in the AER approach, all the synaptic connectivity information is stored in tables used by network routers. Other information may be stored there as well, for example, the strength of the synaptic connection, and the desired delivery delay for the spike.

There may or may not be data associated with each packet, as we will see. No data is necessary with a model that simply conveys a spike. However, a more sophisticated model may deliver a spike rate or a waveform for spikes, through A/D conversion of the output of neuromorphic analog circuits, or could even send continuous waveforms, delivering packets whenever significant changes in voltage occurred.

We will discuss the trade-offs in these connectivity approaches, as well as trade-offs in the neuron modeling, after describing the projects in more detail. There are important differences in scalability, emulation speed, power consumption, and biological accuracy between the connectivity approaches.

2.3 Modeling Plasticity

A static model of the neurons fulfills only some of the requirements for an artificial brain. The other key component is a model of plasticity: how neurons "learn" over time through changes in synaptic sensitivity and through generation of new synaptic connections. Synaptic strength varies in several ways. Presynaptic strength (neurotransmitter availability) is up- or down-regulated (the synapses are facilitated or depressed) through a retrograde process that is not completely understood. Postsynaptic strength is up- or down-regulated through potentiation or depression, by the availability of receptors on the post-synaptic side of the synapse that receive neurotransmitters released on the presynaptic side of the synapse. Post-synaptic strength is modulated by several mechanisms including spike-timing-dependent plasticity (STDP), that increases receptor concentration (synaptic strength) when a

positive post-synaptic potential is followed by a spike generated in the axon hillock, and decreases synaptic strength when the increase in post-synaptic potential is either late with respect to the spiking activity or does not occur at all. Parallel synapses can form at locations where existing synapses are highly active, and synapses can dissolve when activity is absent for some time. Silent synapses that do not respond to presynaptic activity can be awakened via messenger proteins expressed by the neuron's RNA, and new synapses can form over time, possibly due to other protein expression. While the post-synaptic synapse formation is believed usually to occur initially with spine growth as a precursor, followed by the presynaptic growth, there is some evidence that pre-synaptic formation can occur at the same time, or earlier.

The projects we describe assume a limited form of learning, long-term potentiation, and STDP in the brain. They generally implement at the least some form of basic Hebbian learning [20], i.e., when an axon synapsing on a post-synaptic neuron repeatedly takes part in firing the neuron, the synapses on that axon are strengthened. More-complex and more-specific models of plasticity (e.g. STDP) are implemented in some cases. Various more-sophisticated forms of synaptic plasticity have been proposed and studied in neuropsychology. For example, Allport [21] posits that repeated patterns of activity become an auto-associated engram, exciting neurons that are part of the pattern, and inhibiting those that are not. And finally, in addition to strengthening and weakening of synapses, there is evidence in biological neurons, even in mature brains, for the growth of entirely new dendritic spines, dendrites and synapses (e.g., [22, 23]).

Relatively little is written about the plasticity and learning processes used in the projects we cover. However, the learning mechanism is generally encoded in software that can easily be changed, so the projects do offer an opportunity to experiment with various models of learning.

3. Project Summaries

Many projects around the world have aimed at emulating neural networks.[2] In this paper we have attempted to limit our scope to the most advanced and pragmatic approaches to large-scale neural emulation. In particular, we only consider projects intended to scale to millions of neurons, and projects that have fabricated and tested their designs, at least on a small scale, with currently available technologies. Given this scope, although there are innovative, successful projects with more limited scope, due to space and time limitations, we elected to focus on six projects in this paper that have the most ambitious scope and the most demonstrable results:

1. The SpiNNaker [24] project at Manchester University in the U.K.,
2. The Blue Brain[25] project at École Polytechnique Fédérale de Lausanne (EPFL) in Switzerland,
3. The C2S2 SyNAPSE[26 , 27] project at IBM Research in California,
4. The FACETS [28] project at Heidelberg University in Germany,
5. The Neurogrid [29] project at Stanford University in California, and
6. The IFAT [30, 31] and NeuroDyn [32] projects at the University of California at San Diego.

In the following subsections we look at each of these projects in more detail. In the last subsection, we discuss a few related projects, with a focus on emerging technologies.

3.1 SpiNNaker

The SpiNNaker project at Manchester University is based on fabricating many small CPUs on a chip, the cores communicating through a network on-chip and through a network between chips. The principal investigator, Steve Furber, was a co-designer of the ARM 32-bit RISC microprocessor, and a simplified ARM 968 processor is used for the CPUs on the SpiNNaker chips. Each CPU is designed to simulate about 1,000 neurons, communicating spike events to other CPUs through packets on the network.

The SpiNNaker chip is designed to include

- 18 low-power ARM CPUs, each with about 100KB of local RAM used to store its programming and data,
- 128MB of RAM shared by all 18 CPUs through a DMA controller, used to store synaptic weights and other information, and
- An on-chip network and packet router that connects the 18 CPUs and also connects to 6 adjacent SpiNNaker chips, to reach other CPUs.

The routing of packets in SpiNNaker is carefully designed to balance complexity and bandwidth. AER packets are used, as with most of the other projects described here. Routing tables stored in a content-addressable memory tell the router which packets must be routed to which CPUs, whether off-chip or on-chip. The SpiNNaker chips are connected to adjacent SpiNNaker chips in a 2-dimensional toroid mesh network; each chip has 6 network ports, connected to adjacent chips. The router need not know the eventual destination(s) of a packet, it only needs to know which port(s) to send it to. Routing tables are built and maintained by a separate (programmable) background process responsible for connectivity, plasticity, and learning [33].

SpiNNaker is initially using a simple algorithm for neurons based on Eugene Izhikevich's point neuron model 34]. For the purposes of this paper, we analyze SpiNNaker based on that model, although their software-based architecture could support a variety of more sophisticated neural models.

Their point neuron algorithm is programmed into the local memory of each of the SpiNNaker CPUs. Post-synaptic weights for synapses are stored in the SpiNNaker chip's shared memory; the algorithm fetches the

[2] This field is rapidly evolving, so our descriptions reflect a single point in time for each project represented. The reader is cautioned to consult the latest publications for the most accurate information.

corresponding weight into local CPU memory whenever a spike arrives at one of its "synapses," and recomputes neuron action potentials at 1ms simulation intervals, based on Izhikevich's equations. 16-bit fixed-point arithmetic is used for most of the computation, to avoid the need for a floating-point unit and to reduce computation and space costs.

Because spike delivery time in SpiNNaker is designed to be faster than a biological brain (assuming the network and routing delays are adequately controlled), SpiNNaker allows a delay of up to 15ms to be inserted in delivery of AER packets, in order to simulate longer axons. The goal is to allow the globally asynchronous, locally synchronous design to operate similarly to a biological brain possessing the same neural network.

The maximum number of SpiNNaker chips supported by the packet-address structure is 2^{16} (65,000 chips). About a billion neurons could be simulated in this configuration, if the physical chip placement, network, and other constraints do not limit scalability. The group has done some limited simulations to determine when the network and CPUs become saturated [35]. We will further discuss scalability in the last section.

Work continues on the SpiNNaker project. It is expected that a full 65,000-chip configuration with 18-CPU chips will be built some time in 2012.

3.2 Blue Brain

The Blue Brain Project at EPFL in Switzerland uses an IBM Blue Gene supercomputer with 8,000 CPUs to simulate neurons and STDP in software. Henry Markram at EPFL's Brain Mind Institute is the principal investigator. The Blue Brain group constructed a 10,000 neuron model of a neocortical column from the somatosensory cortex of a 2-week-old rat, and simulated it on the Blue Gene supercomputer. The simulation ran about ten times slower than biological neurons.

The modeled cortical column is about .5mm in diameter and about 2.5mm in height. The model is not a map of real connections in any particular rat; the connections are randomly derived based on the percentage connectivity of neurons of different types in different layers of rat cortical columns. However, the model does attempt to account for the 3D morphology of the neurons and cortical column, using about 1 billion triangular compartments for the mesh of 10,000 neurons. A multi-processor adaptation of the NEURON simulation software [36] was run at this fine grain using Hodgkin-Huxley equations, resulting in gigabytes of data for each compartment, and presumably a high level of bio-realism. Timing, e.g. propagation delays along the simulated compartments of an axon, are incorporated into the simulation. Synaptic learning algorithms are also introduced, to provide plasticity. A visual representation of parts of the cortical column can be displayed for the simulation, allowing researchers to focus on particular parts or phases of the simulation in more detail.

The Blue Brain project is unusual in its goal to simulate the ion channels and processes of neurons at this fine-grain compartmental level. Had the project simply used a "point neuron" model integrating incoming spikes, the simulation could have delivered orders of magnitude higher performance, but Markram opted for a higher level of bio-realism.

Of course, software emulation of neurons on large computers, including the bio-realistic fine-grain compartmentalized emulation used in Blue Brain, has been used widely in computational neuroscience laboratories; we mention some other projects at the end of this section. However, we chose to include the Blue Brain project in this paper as the best example of this approach, because of its combination of large scale and bio-realism.

Work on the Blue Brain project is now progressing to a second phase of work. The team cites two new directions: incorporating molecular level processes, and simulating more of the brain through additional parallelism. No publications are yet available on this work, to our knowledge.

3.3 C2S2

Dharmendra Modha's Cognitive Computing Group at IBM Alamaden Research Lab received funding in 2008 from DARPA's SyNAPSE initiative with their proposal "Cognitive Computing via Synaptronics and Supercomputing (C2S2)." Modha has in turn funded professors from 5 universities (Cornell, Columbia, Stanford, Wisconsin Madison, and UC Merced) as part of their project, bringing in expertise in neuroscience, psychology, VLSI, and nanotechnology. We will refer to Modha's project as "C2S2".

Modha's team studied data on biological brains to work toward a "connectome" database of neural connectivity [37], using experimental data from diffusion tensor imaging (DTI) and other techniques. They created a massively parallel cortical simulator called C2, which was initially used at the scale of a rat cortex, and more recently at the scale of a cat cortex, running on IBM's Dawn Blue Gene/P supercomputer, with 147,456 CPUs and 144TB of main memory. In the latter case C2 simulated 1.6B cortical neurons and 9 trillion synapses, using experimentally measured thalamo-cortical connectivity. The simulations incorporated STDP and controlled axon delays.

The C2 simulation used a much simpler model of neurons than the Blue Brain, with single-compartment spiking Iszhikevich-type neurons. As with the Blue Brain, the connectome used did not match the actual connectome of any particular biological brain: it is an approximation based on the tools currently available. However, Modha points out that much can be learned even with these approximations. He reported oscillations in neural firing patterns seen over large areas of the simulated cortex at the alpha and gamma frequencies seen in mammal brains, and groups of neurons exhibited population-specific response latencies matching those in the human cortex.

More recently, Modha has published papers on new "cognitive computing chips" [27], suggesting that IBM research will now turn to hardware for brain emulation. The prototype chip emulates 256 neurons, using a crossbar connecting 1024 input axons to the 256 neurons with

weighted synapses at the junctions. Variations of the chip have been built with 1-bit and 4-bit synapse weights stored in SRAM. Another was built with low leakage to reduce power consumption.

Cross-chip spikes are conveyed asynchronously via AER networking, while the chips themselves operate synchronously. Synapses are simulated using the Izhikevich leaky integrate-and-fire model. The results are identical to the same equations simulated in software, but all 256 neurons on the chip update their membrane voltage in parallel, at 1ms intervals. The details of the AER networking are not specified, so it is not possible to speculate on how that will scale at this time.

3.4 FACETS and BrainscaleS

The FACETS project (Fast Analog Computing with Emergent Transient States) is a consortium of 15 groups in 7 European countries, led by professors Johannes Schemmel and Karlheinz Meier of the Electronic Visions lab at the University of Heidelberg.

In their early work, the "Spikey" neuromorphic ASIC chip was developed. A Spikey chip hosts a total of 128K synapses; it could simulate, for example, 8 neurons with 16K inputs, or 512 neurons with 256 inputs. The goal was to simulate analog neuron waveforms analogous to biological neurons on the same input.

The Spikey neurons communicate with each other digitally, although the neuron circuit is analog. Digital action potentials are routed to synapse drivers, that convert them to voltage pulses that, in turn, control synaptic conductance. The synapse drivers also implement STDP; synaptic weight storage is implemented as static RAM. Synaptic conductance is modulated by an exponential onset and decay.

Whenever an analog neuron circuit reaches an action potential, digital monitoring logic generates a spike event with the event time and the address of the spiking neuron. This event is transmitted on a network to multiple destination neurons that need not be on the same Spikey chip. About 1/3 of the Spikey chip is digital control logic that implements the digital communication between neurons. 16 Spikey chips can be operated on a custom backplane that implements high-speed digital communication between the chips with a fixed and guaranteed latency.

The Spikey chip outputs, inputs, circuit parameters, and neuron interconnections can be monitored and controlled from software running on a host computer. Selected neurons can then be stimulated with experimental spikes, and neuron outputs can be recorded.

More recently, the FACETS researchers developed the HICANN (High Input Count Analog Neural Network) chip and "wafer scale integration" to achieve higher connectivity between simulated neurons. HICANN bears some resemblance to Spikey in that neural emulation is analog, with digital circuits for communication and STDP. However, there are a number of differences. Instead of placing each HICANN chip in a separate package as with Spikey, the entire multi-chip wafer is enclosed in a single sealed package with horizontal and vertical "Layer 1" channels that connect the HICANN chips within and between reticles on a wafer. A total of 352 HICANN chips can be interconnected on the multi-chip wafer, producing 180,000 neurons with a total of 40 million synapses.

Synapses are implemented with groups of DenMem (Dendrite Membrane) circuits. A hybrid analog/digital solution is used for the synapses, and a hybrid of address-encoding and separate signal lines is used for communication. Each DenMem can receive as many as 224 pre-synaptic inputs based on a 6-bit address sent via a Layer 1 channel. The synaptic weight is represented in a 4-bit SRAM with a 4-bit DAC. The post-synaptic signal is encoded as a current pulse proportional to the synapse weight, and can be excitatory or inhibitory. Neuron circuits integrate the DenMem signals. A digital control circuit implements STDP based on temporal correlation between pre- and post-synaptic signals, updating the synaptic weight.

A packet-based "Layer 2" routing protocol is used to communicate between wafers, using pads on the HICANN chips that connect them to the PCB. Layer 2 channels provide 176GB/sec from the wafer to PCB, allowing 44 billion events/second to be communicated between wafers. The Layer 2 wafer-to-wafer channels are handled by FPGAs and OTS switches on the PCB with 1-10 Gbit Ethernet links.

The HICANN chips implement an adaptive exponential integrate and fire (AdExp) model of neurons. This model is somewhat more sophisticated than the standard integrate and fire model used in SpiNNaker, but less sophisticated (and less computational expensive) than Blue Brain's multi-compartmental Hodgkins-Huxley-based model. The FACETS group is now investigating more sophisticated models.

The FACETS neural networks are described in PyNN, a simulator-independent language maintained by neuralensemble.org. PyNN is Python-based and includes operations to create populations of neurons, set their parameter values, inject current, and record spike times. PyNN can be run on a simulator such as NEURON, or can be used on the FACETS host computer to initialize and control the chips. In addition, a neuralensemble.org framework called NeuroTools has been developed to assist in the execution of experiments, and the storage and analysis of results. In recent work [38], software has been developed to automatically translate a PyNN design into a hardware implementation in several stages, optimizing the physical placement of the neural components and connections on HICANN chips.

A follow-on to the FACETS project, BrainscaleS [39], was started in 2011. To date, only high-level directions have been published on BrainscaleS. Two key goals of BrainscaleS are in-vivo recording of biological neural networks and the construction of synthesized cortical networks with similar behavior. The focus is on perceptual systems. The BrainScaleS project is establishing close links with the Blue Brain project and with Brain-i-Nets [40], a consortium producing a set of learning rules based on synaptic plasticity and network reorganization.

3.5 Neurogrid

The Neurogrid project at Kwabena Boahen's "Brains in Silicon" lab at Stanford University uses programmable analog "neurocore" chips. Each 12x14 mm^2 CMOS chip can emulate over 65,000 neurons, and 16 chips are assembled on a circuit board to emulate over a million neurons. The system is built and functional.

Neurogrid uses a two-level simulation model for neurons, in contrast to the point neuron model used in SpiNNaker, and in contrast to the thousands of compartments used in Blue Brain's simulation. Neurogrid uses this approach as a compromise to provide reasonable accuracy without excessive complexity. A quadratic integrate-and-fire model is used for the somatic compartment. Dendritic compartments are modeled with up to four Hodgkin-Huxley channels. Back-propagation of spikes from somatic to dendritic compartments are supported.

Neurogrid uses local analog wiring to minimize the need for digitization for on-chip communication. Spikes rather than voltage levels are propagated to destination synapses. To simplify circuitry, a single synapse circuit models a neuron's entire synapse population of a particular type, and each of these circuits must be one of four different types. The synapse circuit computes the net postsynaptic conductance for that entire population from the input spikes received. Although this approach limits the ability to model varying synaptic strength, and it does not model synaptic plasticity, it greatly reduces circuit complexity and size.

Like SpiNNaker, Neurogrid uses an AER packet network to communicate between-chip spikes. Unlike SpiNNaker's grid organization, Neurogrid's chips are interconnected in a binary tree with links supporting about 80M spikes/second (this is a change from earlier work [41] in which Boahen used a grid network). Routing information is stored in RAM in each router. This AER-based networking is referred to as "softwire" connections.

To reduce communication overhead, a single inter-chip spike can target multiple neurons on the destination chip. The postsynaptic input triggered in a target neuron can be propagated to neighboring neurons with a programmable space-constant decay. This requires only nearest-neighbor connections: the synaptic potentials superimpose on a single resistive network to produce the net input delivered to each neuron. A single cross-chip spike can thus reach a hundred neurons. This is analogous to cortical axons that travel for some distance and then connect to a number of neurons in local patch arbors in another cortical column.

Unlike FACETS, which is designed to run orders of magnitude faster than biological neurons, the Neurogrid neuron array is designed to run in real-time. This means that a single AER link can easily service all of the cross-chip spikes for 65,000 neurons. Furthermore, the on-chip analog connections can easily service their bandwidth, and it seems likely that the binary routing tree connecting the 16 Neurogrid chips on a circuit board can easily support a million neurons. Thus, the only potential bottleneck for Neurogrid might be in routing between multiple boards in the future.

Like FACETS, the neurocore chips are programmable. Each neurocore models the ion-channel behavior and synaptic connectivity of a particular neuron cell type or cortical layer. The Neurogrid neuron circuit consists of about 300 transistors modeling the components of the cell, with a total of 61 graded and 18 binary programmable parameters. Synapses can be excitatory, inhibitory, or shunting. The Neurogrid group has demonstrated that their neurons can emulate a wide range of behaviors.

The Neurogrid team has encouraged others to build on their work, teaching courses training students to build neural networks on their framework, and making their silicon compiler available to allow others to design neuromorphic systems for fabrication. The descriptions are written in Python.

3.6 IFAT and NeuroDyn

Like the Neurogrid and FACETS projects, Gert Cauwenberghs and colleagues at the Institute for Neural Computation (INC) at the University of California at San Diego chose to use analog neuromorphic circuit chips to model neurons. They have produced two different chips, IFAT and NeuroDyn, with different goals.

The initial IFAT (Integrate and Fire Array Transceiver) chip, built in 2004, could emulate 2400 simple neurons. A separate microcontroller on the same circuit board used analog-digital converters and an AER lookup table to deliver spikes to the IFAT chips based on a global "clock cycle." The INC group applied the IFAT chips to various applications, including Laplacian filters to isolate vertical edges on images, and spatiotemporal filters to process a spike train from an artificial retina, constructing velocity-selective cells similar to those found in the medial-temporal cortex in the human brain, demonstrating brain processing.

The latest version of the IFAT chip emulates 65,000 neurons. The new system, called HiAER-IFAT (Hierarchical AER IFAT), uses a tree of routers for delivery of AER events [42]. The tree is built using Xilinx Spartan-6 FPGAs connecting to the IFAT chips at the leaves. HiAER-IFAT has been demonstrated with 250,000 neurons. Like SpiNNaker, all of the connectivity information is held in RAM in the routing tables of the intermediate nodes, in this case the non-leaf nodes of a hierarchy. Unlike SpiNNaker, the maximum number of routing "hops" is logarithmic in the number of neurons. However, it is possible that the HiAER-IFAT routers in the highest level of the hierarchy could become overloaded if there is insufficient locality of reference.

The INC group has also designed a "NeuroDyn" chip, which is the most sophisticated of all of the neuromorphic chips discussed in this paper, in terms of bio-realism and neuron emulation. Their neuron emulation supports 384 parameters in 24 channel variables for a complex Hodgkin-Huxley model. This level of emulation is important, for example, in examining the effects of neuromodulators, neurotoxins, and neurodegenerative diseases on ion channel kinetics. However, NeuroDyn is not designed for large-scale brain emulation: each chip emulates only 4 neurons and 12 synapses.

In contrast to IFAT and all the other projects that generate discrete spike events to be delivered by AER or other means, NeuroDyn emulates neural and synaptic dynamics on a continuous basis. Matlab software on a workstation can monitor and control each neuron's membrane potential and channel variables, and can adjust the 384 NeuroDyn emulation parameters to tune to any desired neuron behavior. The parameters are stored on chip in digital registers. Experiments analogous to patch-clamping biological neurons can be performed on NeuroDyn neurons through the software.

3.7 Other projects

Some other projects are worth mention because they address the challenges of an artificial brain in novel ways, although they have not yet progressed enough to include in our comparison at this time. Additional projects are also surveyed in papers by de Garis et al [43], although a number of those projects are aimed at higher-level models of the brain, not the direct emulations surveyed here.

The BioRC [44] project at the University of Southern California, led by the second co-author of this paper, is worth mention because of its radically different technology approach: artificial neurons and connections are proposed to be built from carbon nanotubes and other nanodevices like nanowires or graphene transistors. The long-term goal of the BioRC research project is the development of a technology and demonstration of electronic circuits that can lead to a synthetic cortex or to prosthetic devices. However, this project is still at an experimental stage, designing individual neurons and small neural networks, so we did not include it in our comparison.

The BioRC project aims to meet all of the challenges discussed earlier in this paper, and the choice of emerging nanotechnologies is posited to be required in order to achieve all the challenges posed. While experiments to date have involved carbon nanotube FET transistors, other nanotechnologies are under investigation. Carbon nanotubes have some distinct advantages, not provoking an immune system reaction or corroding in contact with living tissue, as well as the obvious advantages of being extremely small (a few nm in diameter) and low power. Finally nanotechnologies like carbon nanotubes offer a possible future advantage if they can be arranged and rearranged into 3-D structures of transistors and circuits to support the connectivity and structural plasticity challenges faced when building an artificial brain.

The BioRC neural circuits can be arranged to implement neurons with many variations in structure and behavior. The neural circuits are also designed with inputs that act as "control knobs" to vary neural behavior. The control knobs can be used to create neurons with differing characteristics (e.g. spiking vs. bursting), or can be used as control inputs representing external influence on neural behavior (e.g. neurohormones). A variety of synapse circuits, axon hillocks, and dendritic arbors have been designed to illustrate temporal and spatial summation, STDP, dendritic computations, dendritic spiking, dendritic plasticity, and spiking timing variations. A CMOS chip containing many of the circuits has been fabricated. Finally, a single synapse with a carbon nanotube transistor has been constructed and tested in collaboration with Chongwu Zhou [45].

BioRC's neural models, especially the synaptic models, are more complex than most of the major projects, with the exception of Markram's Blue Brain. Interconnections in existing CMOS technology are believed to be the primary challenge to whole brain emulation for this project, although newer flip-chip technologies can ameliorate connectivity problems significantly. Self assembly with nanodevice transistors, like that performed by Patwardhan *et al.* [46] shows promise for future whole brain emulation with analog circuits.

Memristors are another nanotechnology being implemented in neural circuits, with the pioneering work at HP, where the first fabricated memristors were invented [47].

In addition, various other research groups have made progress towards more advanced neural simulations:

- Eugene Izhikevich, CEO of the Brain Corporation, together with Nobel prize winner Gerald Edelman, simulated a million spiking neurons and 500 million synapses tuned to approximate recorded in-vitro rat cortex neural responses [48]. Their neural model was slightly more sophisticated than the one used in Modha's simulations, separating the soma from multiple dendritic compartments. Like Modha, they found that waves and rhythms emerged. They also found their simulation highly sensitive to the addition or removal of a single neural spike.
- Giacomo Indiveri's Neuromorphic Cognitive Systems Lab at the Institute of Neuroinformatics at the University of Zurich have built biomimetic hybrid analog / digital CMOS VLSI chips for specific functions such as real-time sound recognition and optic flow sensors, using quite detailed neuronal models [49].
- The Computational Neurobiology Lab at Salk Institute as well as the INC lab at UCSD perform detailed software neuron simulation and advanced recording of biological brains, for example to model learning [50]. The MCell project simulates detailed diffusion machinery and other biomolecular processes at synapses.
- Farquhar and Hasler at Georgia Tech describe a programmable neural array [51],with analog neural circuits.

4. Analysis and Comparisons

Each of the projects we discuss address some challenges to artificial brain construction directly. However, none of the projects masters all of them. In this section of our paper, we examine four challenges:

1. Bio-realism of the neural computation model, i.e., the project's ability to emulate the behavior of biological neurons,

2. Bio-realism in neural connectivity, including fan-in and fan-out,
3. Bio-realism in synaptic and structural plasticity, i.e., whether an artificial brain will learn and adapt like a biological brain, and
4. Scalability of all the above, including power and space requirements, for tens of billions of neurons and hundreds of trillions of connections.

4.1 Neuron Emulation

The projects we focused on in this paper use three different technological approaches to the emulation of neurons. SpiNNaker uses a "neuroprocessor" approach, emulating neurons in software on loosely-coupled (networked) CPUs. Blue Brain and the original C2S2 work use a "neurosimulation" approach, emulating neurons in software on tightly-coupled (shared memory) CPUs in a supercomputer. FACETs, Neurogrid, and NeuroDyn use a "neuromorphic" approach, with analog circuitry for neural computations.

Independent of the technological approach, the projects differ substantially in the level of bio-realism and computational sophistication in their emulation of neurons and synapses:

1. The simplest approach is the *point neuron* model, as recommended by Izhikevich, in which a neuron's synaptic inputs enter into differential equations to compute the output of the neuron over discrete time intervals. SpiNNaker and the C2S2 work have used such a model.
2. A point neuron model implemented in analog circuitry is potentially more sophisticated, depending on the complexity of the circuit, since the circuit can perform continuous real-time integration of signals in contrast to the discrete-time intervals used in software emulations. The NeuroDyn chip implements a particularly sophisticated point neuron, with hundreds of parameters.
3. A *two-level* analog model such as Neurogrid's two compartments, or the FACETs HICANN chip's separate dendritic membrane circuits, allows more sophisticated neural emulations, depending on the complexity of the compartment emulations.
4. The most bio-realistic approach among the projects is Blue Brain's *fully compartmentalized* model of the neuron, representing a biological neuron as hundreds of independent compartments, each producing an output based on adjacent ion channels and regions. These result in an integrated neural output at the axon hillock compartment, but also allow for local dendritic spikes and back-propagation of action potentials to dendrites. Blue Brain uses computationally expensive Hodgkin-Huxley equations to compute the potential bio-realistically in each compartment.

The neuromorphic approach avoids the substantial computational overhead of software simulation, and may produce a more biologically-accurate result in less time than point neuron software simulations using Izhikevich's equations. On the other hand, while neuromorphic analog circuits can produce results many orders of magnitude faster than real neurons or a software simulation like Blue Brain, there is still a remaining question about whether their fixed neuronal structure adequately captures biological neuronal behavior.

Because the connectome used in the Blue Brain simulations is not identical to any biological brain, it is difficult to observe identifiable functional behavior from the cortical column they simulate, except in very abstract ways. Since none of the systems can be directly compared to biological brains, it remains an open question what neural complexity is required to demonstrate biological behavior.

Keep in mind that biological neurons are slow in comparison to current electronics, differing by at least a factor of 10^6 in speed, if we compare the speed of a simple logic gate with the speed of a neuron. However, it takes many machine cycles and tens of thousands of gates executing software instructions in order to simulate a neuron. There is also significant overhead due to communication between processors, further slowing execution. At one point, the Blue Brain neurons were about ten times slower than biological neurons in simulations, and used about 8000 processors to simulate 10,000 neurons in a cortical column of a rat. This highlights the need for massive parallelism, and the performance degradation when simulation is performed on serial processors.

Note that the artificial brain projects can be grouped into two overall categories: simulating neurons with digital hardware (Blue Brain, C2S2, and SpiNNaker), or simulating neurons with in analog hardware (FACETs, NeuroDyn, and Neurogrid). Most projects seem to rest at the extremes of processing variations: massive multiprocessor software simulations or analog neuromorphic circuit emulations. One could speculate that special-purpose digital hardware built with FPGAs or as ASICs would explode in size and complexity due to the non-linear additions and multiplications occurring in the dendritic arbor, forcing digital implementations that implement the dendritic arbor for each neuron to be significantly simplified over software implementations. Because of the relative simplicity of analog computations compared to digital computations, most hardware approaches have exploited the ability to manipulate currents and voltages by means of analog electronics, inspired by the seminal work of Misha Mahowald [19] and Carver Mead [52]. While the analog computations are inexact and subject to fabrication and environmental variations, biological neural networks exhibit variability in behavior as well, and still perform well under differing circumstances.

The troubling thought is that there are no definitive results to indicate how detailed the model of the brain and neurons must be in order to demonstrate intelligence. Non-linear dendritic computations and dendritic spiking are shown to occur in the biological brain (e.g., by Polsky [7]), but perhaps such biological structures could be supplanted with more intricate connectivity between simpler neuronal

structures in an artificial brain, much as the analogy between Turing machines and modern supercomputers, with elaborate programming schemes in a Turing machine replacing software running on a more-complicated execution engine. Thus, while some attempts are less bio-realistic in their models of neuronal computations, they might be able to demonstrate equivalent intelligence with added sophistication of connectivity over other models.

4.2 Synaptic Connectivity

As with the neuron models, there are a number of technical approaches to modeling synaptic connectivity. Blue Brain uses software calls. The others generally use a digital networking approach, but Neurogrid, FACETS, NeuroDyn, and IFAT use direct wiring for short distances.

As with neuron emulation, and independent of technical approach, the important differences in connectivity approaches are the resulting bio-realism and capabilities:

1. Delivered content: the synaptic connectivity model may simply deliver a spike as an event to another neuron, the entire spike voltage waveform may be delivered, or there may be continuous connection, at least at some small time granularity, so that sub-threshold voltages can affect a synapse. There are key questions for neuroscience to answer, here, before we can judge what must be delivered. There is some evidence that implementing the spikes as events alone is adequate, and this would vastly simplify the circuits and technology to emulate synaptic connectivity, but there is dissention concerning this assumption.
2. *Connection distance: the projects differ in their* ability to deliver output to distant vs. nearby neuron synapses, or in the properties of their short vs. long connections, e.g. with direct connections versus AER packets.
3. Connection delays: biological axon/synapse connections differ in their time delays, particularly for axons and dendrites that synapse over longer distances, or through the thalamus. A model that treats all connections the same cannot model this. However, it may be possible to insert delays to simulate longer connections in all of the projects, and even with direct wiring, delay circuits could be inserted. AER packet delays can be unpredictable, but a variable delay can be inserted to achieve any desired delivery time, if AER delays can be adequately bounded. Of course, tracking and correcting delivery times add complexity to those systems.
4. Fan-in/fan-out: There are limitations in the fan-in and fan-out allowed by all the technologies, and there will be bigger delays with larger fan-in and fan-out, with either direct connections or with AER packet delivery. We will examine connectivity scalability in Section 4.4.
5. Timing and other issues: A final challenge related to connectivity is the neural sensitivity to spike-arrival timing at the synapse. Late spikes result in synaptic depression in biological synapses. Arrival of spikes in a predictable manner supports rate coding, believed to be a mechanism that conveys more information than a more-binary interpretation of spikes with spike/no-spike processing. Thus, connecting the brain physically is a major challenge, but predictable spike-arrival timing further complicates the connectivity problem enormously. In addition, communication between proximal neurons occurs via astrocytes as well and is postulated to occur via electromagnetic waves and other signals, further complicating the wiring.

An architecture with synchronous delivery of spikes introduces a timing issue. For example, connectivity in the original IFAT chip was based on delivering spikes on global clock cycle intervals, with neurons computing their output state on each cycle, while HiAER-IFAT provides asynchronous operation.

Continuous analog connectivity works well for short distances, as demonstrated by a number of the projects, but direct wiring to all neurons at the scale of an artificial brain requires massive connectivity not yet possible in modern digital circuits. The brain does seem to follow a Rent's rule [11] just as digital systems do, in that there is a relationship between the number of connections emerging from a volume of brain tissue compared to the size of the brain tissue enclosed. However, all modern digital systems exploit some form of multiplexing for communications at any distance. An artificial brain that did not multiplex connections, sharing wires between distant parts of the brain, would likely be unable to support 10,000 connections per neuron using current technologies.

Thus, all of the projects have a challenge with the brain's dense synapse fan-in and axon fan-out. To date, the solution of choice is AER packet networking. In the case of neuromorphic analog circuits (i.e., for all but SpiNNaker and Blue Brain), the use of AER requires some form of A/D and D/A conversion, or at least detection and delivery of a spike threshold voltage or spike event. This is problematic, because the per-neuron circuitry required for conversion and for the routing and decoding logic for the very large address space (37 bits in the worst case) is much larger than the neuron emulation circuit itself, perhaps orders of magnitude larger. Recent evidence points to significant connectivity outside each individual cortical column, implying many non-local connections [11].

In SpiNNaker's "neuroprocessor" approach, there is no A/D conversion and the cost of the network routing logic is amortized over 1,000 emulated neurons per CPU. However, both SpiNNaker and the neuromorphic analog circuits face another problem with AER networking: routing packets in real-time from tens of billions of neurons is a major challenge. We will examine networking scalability in Section 4.4.

Another issue with AER networking is the timing of spikes. Neurons adapt to early and late signals over time, and some signal timing tuning is performed by the oligodendrocytes that form a myelin sheath on the neurons'

axons. In an emulated brain, with packet communication over a network, timing of emulated spikes originating from the same neuron is uncertain from second to second. However, proper synchronization can be achieved by inserting delays or inserting delays or reserving bandwidth [32, 53].

4.3 Plasticity and Learning

Plasticity and learning present one of the biggest challenges to artificial brain projects, partly because we don't fully understand how these work in the brain, partly because such biological mechanisms are quite complex (e.g. Kandel's seminal research on memory [54]), and partly because our technologies are far less plastic than neural tissue. We have experimental evidence for some basic synaptic plasticity mechanisms, but knowledge about plasticity, learning, and memory is far from complete. Basic Hebbian learning could likely be the "tip of the iceberg". There is already evidence that Hebbian learning applies not just at the level of individual synapses, but to groups of interacting neurons, and many forms of learning may have evolved in the brain, or in different parts of the brain [22]. There is also evidence for neurons growing new dendrites and synapses to create new connections as well as changing the "weight" of existing synapses by increasing or decreasing the number of neurotransmitter vesicles or receptors for the neurotransmitters. Finally, learning can also occur through neurons dying or being created.

The projects have generally not specified the mechanisms for learning in an artificial brain. However, they generally explain how learning would be programmed, once a learning algorithm is specified. There are three approaches:

1. In an all-software solution like Blue Brain, the learning algorithm is of course implemented in software, which makes it easy to change. Connections can added and removed, their synaptic strengths can be changed, new neurons can be created, and existing ones can be removed.
2. A separate software module can be used for learning with a solution that implements synapses through AER packets, and which uses tables to store both the connection information and the "weight" of each connection. Connections can be created and destroyed in the table as well. This is the case for SpiNNaker's tables, the tree of connections in FACETs and Neurogrid, and the separate tables used in IFAT. It should also be possible to create and remove neurons via the separate process, given an appropriate interface to the neuron implementation.
3. In a solution with direct wiring, at least some of the learning mechanism must be implemented in the artificial neuron itself. FACETS includes hardware-based STDP in the HICANN chips. However, a separate mechanism for the creation and removal of neurons would be needed in this case, and there is currently no practical solution for the growth of new synaptic connections as direct "wires," particularly over longer distances.

Although neuroscience's understanding of learning and plasticity is currently limited, and artificial brain projects have offered incomplete solutions to date, these projects will likely prove important in our understanding going forward. It is fortunate that most of the brain emulation projects have left the learning mechanism open, to be programmed in software. It is very difficult to understand synaptic changes and learning in a heavily-interconnected biological brain: experiments *in vivo* are very difficult, and experiments *in vitro* do not deal with sufficiently-complex neuron networks. In an artificial brain, in contrast, it is possible to experiment with many different plasticity mechanisms, monitoring all the neuron connections over time, and observing the learning behavior. These experiments could prove to be the most useful contribution of artificial brain projects over the coming decade.

In addition to the learning problem, an unsolved problem is the initial configuration of an artificial brain prior to developmental change and learning. Embryologically, neurons are of many different types and have at least some genetically-determined connections or connection patterns that are almost certainly essential to intelligent behavior, and without which learning would probably not occur. It is known that connections are pruned as the brain matures and knowledge specializes the neural circuits. Even at the individual neuron level, the initial configuration of synaptic strengths, threshold voltages, and integration behavior in an artificial brain, along with models of learning mechanisms to be used, will certainly determine whether learning occurs later. In Ishikevich's favored learning model, for example, no initial values for the parameters a, b, c, and d are specified for his equations. An open question is what should these be.

4.4 Overall Scalability

Armed with some understanding of the technologies and mechanisms proposed for artificial brains, we can now examine practical issues of overall scalability.

All-software solutions such as the Blue Brain project do not directly address scalability. Instead, the problem is reduced to finding a sufficiently large supercomputer to run the software at the desired scale. Since that project is already experiencing limitations running on one of the largest supercomputers available to date, and they are emulating less than .0000001% of the neurons in the human cortex, there are obviously scaling issues here, and there are power and scaling constraints on the largest supercomputers that can be built. Another alternative would be a software approach decomposed into processes that could run on many distributed computers, e.g. using computing power on many sites or cloud computing. We are not aware of a solution using this approach to date, but the communication latency and bandwidth could prove to be a problem with this approach, even if enough processing power could be secured.

In all the other projects, a hardware solution is proposed. The scaling challenges in these projects fall into four categories:

- *Physical size and packaging:* Each project proposes integrated circuit chips that emulate some number of neurons and connections. For example, the NeuroDyn chip can perform Hodgkin-Huxley simulations of 4 neurons with limited fan-in, the Spikey chips can simulate 8 neurons with fan-in comparable to cortical neurons, and the current SpiNNaker chip is projected to perform about 18,000 much simpler point-neuron simulations. To scale to the billions of neurons in the mammalian cortex, millions of chips would be required using present-day CMOS technology, even with the simple point-neuron model.
- *Connectivity:* A separate problem involves the fan-in and fan-out of connections between the emulated neurons within and between the chips. No current technology allows for reasonably-sized emulated neuron circuits with an average of 10,000 inputs and outputs to other neuron circuits on a chip: the combinatorial explosion of input wires to each neuron would overwhelm any integrated circuit layout with more than a few dozen neurons. Thus, almost all of the projects we detailed have turned to digital networking. However, there are bandwidth and circuitry limitations to networking, as we will discuss shortly.
- *Power and heat issues:* The human brain is incredibly power-efficient. It is a 3-dimensional computer with liquid coolant/power delivery. Even with the most energy-efficient integrated circuit technologies we have, heat dissipation and total power requirements will be a problem in scaling to the size of biological brains. The neuromorphic solutions (FACETS, Neurogrid, NeuroDyn, IFAT) are most promising in terms of lowest power consumption per emulated neural time unit. For example, the HICANN wafer uses only 1 nJ of energy per synaptic transmission, less than a single instruction on an ARM processor. At the same time, the speed of neural emulation and communication in neuromorphic solutions can run 10,000 times faster than biological equivalents.
- *Ancillary issues:* The artificial brain projects propose ancillary mechanisms that must also scale along with the artificial brain. For example, if a software process separate from the actual neuron implementations is responsible for learning, then it must scale alongside. A question arises as to whether the learning process be decomposed into many cooperating learning processes on multiple processors distributed throughout a network.

How well does AER networking scale? If there are about 40 billion neurons in the human cortex and thalamus, with an average axon fan-out to 10,000 synapses, firing at an average of 10 times per second, then AER networking would need to deliver about $4x10^{14}$ packets per second, with at least $4x10^{10}$ originating packets per second. To put this in perspective, it is recently estimated that the total U.S. user Internet traffic averages about $8x10^{8}$ packets per second. Admittedly, the AER packets are fewer bytes, and over short distances, but the routing overhead is comparable, and the routing tables are much bigger, given 40 billion destinations. Even if the firing rate is significantly lower than our estimate, the total traffic is staggering when taken as a whole.

Luckily, the actual number of inter-chip packets might be much smaller. There is evidence that interconnectivity is much higher within a cortical column and minicolumn. With an estimated 10,000 neurons in a cortical column, cortical columns could fit entirely on planned second-generation chips for projects such as SpiNNaker and FACETS. If 99% of connectivity is within a column, this reduces the inter-column and inter-chip bandwidth 100 times, and earlier-mentioned research on a "Rent exponent" by Bassett *et al.* [11] suggests that locality of connectivity may extend beyond cortical columns.

Even without locality of reference assumptions, the SpiNNaker group provides some evidence that their AER network can scale to 1 billion neurons with a fanout of 1,000 synapses per neuron [55]. While falling short of the scale of the human cortex, this is a promising result. Their torus of 65,000 interconnected chips, each with its own router, and each connected to 6 neighbors, allows more even distribution of load than hierarchical networks and subnetworks. Thus with the right networking topology, some help from locality of reference, and a significant portion of computing power dedicated to routing, it may be possible to support AER networking on the scale of the human brain, but this remains to be demonstrated.

Turning our attention from neuron connectivity to neuron emulation, there are scalability issues there as well.

As just discussed, Blue Brain emulates a small, fractional percent of brain neurons in much less than real time.

The analog neuromorphic circuit approach requires less hardware due to the special-purpose nature of the circuits, and the economy of analog "computations," absent any interconnection hardware. Because the computations are inexact, direct comparison is not possible. However, neurons with complexity somewhere between Blue Brain and SpiNNaker could be realized with synaptic plasticity and dendritic computations with less than a million transistors per neuron [38]. Given the potential for over a billion transistors per chip with current technology, each neuromorphic chip could hold over 1,000 neurons. However, note that many millions of chips would still be required for an artificial brain, and the connectivity and structural plasticity problems with neuromorphic analog circuits remain. Major breakthroughs in nanotechnology that allow 3-dimensional construction and real-time modification of electronic circuits would be required to achieve an analog whole brain.

The highest scale is achieved by SpiNNaker, with its simpler neuron model. However, even using SpiNNaker chips with 18 CPUs, over a million chips would be required for the human cortex. And for bio-realism of the complexity of Hodgkin-Huxley with two or more levels per neuron, and synapses and dendritic arbors with commensurate complexity, a much smaller number of neurons could be emulated by each CPU.

In summary, SpiNNaker's neuroprocessor approach gives the highest scalability but with limited bio-realism, the neurosimulation approach gives the highest bio-realism with scalability limited by the largest supercomputer available, and the neuromorphic approaches are in between in bio-realism, being limited in scalability until the "wiring" and circuit density problems are solved.

5. Conclusions

We are a long way from a working artificial brain. Given our limited of understanding of biological neurons and learning processes, the connectivity scalability issues, and the substantial computing power required to emulate neuronal function, readers may understandably be skeptical that a group of interconnected artificial neurons would possibly behave in a fashion similar to the simplest animal brains, let alone display intelligence.

With the drawbacks of all three approaches we discussed, it seems that there is not one good approach to all of the problems. In the near term, software is easier to experiment with than hardware, and connectivity and structural plasticity are practical to attain in partial brain emulations. Experiments with software may produce more progress on requirements for neural modeling. At the same time, experiments with analog hardware might demonstrate economies of scale and use of nanotechnologies for future whole brain emulation.

In our opinion, the most promising approaches depend on the goals and timeframe:

1. In the short term, scalability to the size of a mammalian brain is not practical, but software simulation seems most promising for emulation and understanding of small networks of neurons, e.g. to experiment with learning algorithms, since software is so easily modified.
2. An approach like SpiNNaker's, with loosely-coupled processors and AER networking, seems most likely to yield neural emulation on the scale of the entire brain in the medium term. The result might or might not behave like a biological brain, given the simplifying assumptions, e.g. using a point neuron model and delivering spikes rather than continuous synaptic connectivity.
3. In the long term, if a solution such as DNA-guided self-assembly of nanoelectronics is developed to allow self-guided wiring and re-wiring of dendrites and axons, a neuromorphic analog solution seems the only approach that can give bio-realism on a large scale.

Neuroscientists are uncovering additional mechanisms and aspects of neural behavior daily, and neural models for artificial brains may become significantly more complex with future discoveries. However, an artificial brain, limited and simplified, does provide a test bed for research into learning and memory, and we expect that substantial progress will be made through new technologies and ideas over the coming decades. In addition, research on smaller-scale artificial neural networks still provides considerable value in applications such as signal processing and pattern recognition, and experiments with neural networks may give us insights into learning and other aspects of the brain.

Independent of any value to neuroscience, efforts on artificial brain emulation also provide value to computer science. Projects such as SpiNNaker yield new computer architectures that have other valuable applications, and emulation of the brain may yield new approaches to artificial intelligence. If the goal is artificial intelligence or new computer architectures rather than bio-realistic brain emulation, then it is also possible that simpler neuron emulations would be adequate. Thus, we see value in continued research despite our pessimism about the timeframe and current technologies for brain emulation.

Acknowledgment

We would like to thank Kwabena Boahen, Gert Cauwenberghs, Luke Muehlhauser, Johannes Schemmel, and Thomas Sharp for their inputs on this paper. However, we are solely responsible for any remaining errors.

References

[1] National Academy of Engineering (nae.edu), Grand Challenges for Engineering, www.engineeringchallenges.org, 2010.

[2] W.McColloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", in Bulletin of Mathematical Biophysics Vol 5, pp 115–133, 1943.

[3] F. Rosenblatt, "A Probabilistic Model for Information Storage and Organization in the Brain," Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386–408, 1958.

[4] A.L. Hodgkin, A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerves," Journal of Physiology 117, pp 500–544, London, 1952.

[5] W. Rall, "Branching dendritic trees and motoneuron membrane resistivity," Experimental Neurology 1, pp 491–527, 1959.

[6] T. Berger, et al, "A cortical neural prosthesis for restoring and enhancing memory," Journal of Neural Engineering 8, 4, August 2011.

[7] Shepherd, G. "Introduction to Synaptic Circuits," in The Synaptic Organization of the Brain, edited by Gordon Shepherd, 5th edition, Oxford University Press, 2004.

[8] A. Polsky, B. W. Mel, J. Schiller, "Computational subunits in thin dendrites of pyramidal cells, Nature Neuroscience, www.nature.com/natureneuroscience, 2004.

[9] A. Losonczy, J. K. Makara, and J. C. Magee, "Compartmentalized dendritic plasticity and input feature storage in neurons," Nature, vol. 452, pp. 436–441, March 2008.

[10] S. Remy, J. Csicsvari, and H. Beck, "Activity-dependent control of neuronal output by local and global dendritic spike attenuation.," Neuron, vol. 61, pp. 906–916, March 2009.

[11] S. Bassett, D.L. Greenfield, A. Meyer-Landenberg, D. R. Weingerge, S.W.More, E.T.Bullmore, "Efficient Physical Embedding of Topologically Complex Information Processing Networks in Brains and Computer Circuits, PLoS Computational Biology 6, 4, 2010.

[12] R. D. Fields, The Other Brain: From Dementia to Schizophrenia, How New Discoveries about the Brain Are Revolutionizing Medicine and Science. Simon & Schuster, 1 ed., December 2009.

[13] J. Joshi, A. Parker, K Tseng, "An in-silico glial microdomain to invoke excitability in cortical neural networks," IEEE International Symposium on Circuits and Systems, Rio de Janeiro, Brazil, May 2011.

[14] A. Sandberg, N. Bostrom, Whole Brain Emulation: A Roadmap, Technical Reprot 2008-3, Future of Humanity Institute, Oxford University, 2008.

[15] H. de Garis, C. Chuo, B. Goertzel, L. Ruiting, "A world survey of artificial brain projets, Part I: Large-scale brain simulations," Neurocomputing 74, Issues 1-3, pp. 3-29, December 2010.

[16] B. Goertzel, R. Lian, I. Arel, H. de Garis, S. Chen, "A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures," Neurocomputing 74, Issues 1-3, pp 30-49, September 2010.
[17] C. Koch, I. Segev, Methods in Neuronal Modeling, 2nd Edition, Massachusetts Institute of Technology, Cambridge, MA, 1998.
[18] Izhikevich, E. M., "Which model to use for cortical spiking neurons?," IEEE Trans. Neural Networks 15, 2004, pp 1063-1070.
[19] M. Mahowald, Computation and Neural Systems, PhD thesis, California Institute of Technology, 1992.
[20] D. Hebb, The Organization of Behavior, Wiley & Sons, New York, 1949.
[21] D. Allport, "Distributed memory, modular systems, and dysphasia," in S. Newman and R. Epstein, Current Perspectives in Dysphasia, Churchill Livingstone, Edinburgh, 1985.
[22] N. Ziv, "Principles of glutamatergic synapse formation: seeing the forest for the trees," Current Opinion in Neurobiology, vol. 11, pp. 536–543, October 2001.
[23] C. Verderio, S. Coco, E. Pravettoni, A. Bacci, and M. Matteoli, "Synaptogenesis in hippocampal cultures," Cellular and Molecular Life Sciences, vol. 55, pp. 1448– 1462, August 1999.
[24] S. Furber and S. Temple, "Neural Systems Engineering", in Studies in Computational Intelligence, Springer Verlag, 2008.
[25] H. Markram, "The Blue Brain Project", in Nature Reviews, Volume 7, February 2006, pp 153-160.
[26] D. Modha, et al, "Cognitive Computing: Unite neuroscience, supercomputing, and nanotechnology to discover, demonstrate, and deliver the brain's core algorithms," Communications of the ACM 54, 8, pp 62-71, August 2011.
[27] P. Merolla, J. Arthur, F. Akopyan, I. Nabil, R. Manohar, D. Modha: "A Digital Neurosynaptic Core Using Embedded Crossbar Memory with 45pJ per Spike in 45nm", IEEE Custom Integrated Circuits Conference (CICC), San Jose, 2011.
[28] J. Schemmel et al, "A Wafer-Scale Neuromorphic Hardware System for Large-Scale Neuron Modeling", in Proceedings of the 2010 IEEE International Symposium on Circuits and Systems, 2010.
[29] R. Silver, K. Boahen, S. Grillner, N. Kopell, K. Olsen, "Neurotech for Neuroscience: Unifying Concepts, Organizing Principles, and Emerging Tools", Journal of Neuroscience, pp 11807-11819, October 2007.
[30] R. Jacob Vogelstein et al, "A Multichip Neuromorphic System for Spike-Based Visual Information Processing", in Neural computation, Volume 19, 2007, pp 2281-2300.
[31] J. Park, T. Yu, C. Maier, S. Joshi, G. Cauwenberghs, "Hierarchical Address-Event Routing Architecture for Reconfigurable Large Scale Neuromorphic Systems," IEEE International Symposium on Circuits and Systems, Rio de Janiero, May 2011.
[32] T. Yu and G. Cauwenberghs, "Analog VLSI Biophysical Neurons and Synapses with Programmable Membrane Channel Kinetics", IEEE Transactions on Biomedical circuits and Systems, 4, 3, p 139-148, June 2010.
[33] X. Jin, A. Rast, F. Galluppi, S. Davies, S. Furber, "Implementing Spike-Timing-Dependent Plasticity on SpiNNaker Neuromorphic Hardware," IEEE World Congress on Computational Intelligence, Barcelona, Spain, July 2010.
[34] E.M. Izhikevich, "Simple Model of Spiking Neurons," IEEE Trans. Neural Networks, vol. 14, no. 6, 2003, pp. 1569–1572.
[35] A. Rast, S. Furber, et al, "Scalable Event-Driven Native Parallel Processing: The SpiNNaker Neuromimetic System," CF '10, Bertinoro, Italy, May 2010.
[36] NEURON simulation software, http://www.neuron.yale.edu/neuron, 2005.
[37] D. Modha and R. Singh, "Network architecture of the long-distance pathways in the macaque brain," Proceedings of the National Academy of Sciences of the USA 107, 30 (June 2010), 13485–13490.
[38] M. Ehrlich, et al, "A software framework for mapping neural networks to a wafer-scale neuromorphic hardware system," Proceedings Artificial Neural Networks and Intelligent Information Processing Conference, 2010.
[39] BrainScaleS Project Overview, http://brainscales.kip.uni-heidelberg.de.
[40] Brain-i-Nets, consortium overview, http://brain-i-nets.kip.uni-heidelberg.de/.
[41] J Lin, P Merolla, J Arthur and K Boahen, "Programmable Connections in Neuromorphic Grids," 49th IEEE Midwest Symposium on Circuits and Systems, pp 80-84, IEEE Press, 2006.
[42] S. Joshi, S. Deiss, M. Arnold, J. Park, T. Yu, G. Cauwenberghs, "Scalable Event Routing in Hierachical Neural Array Architecture with Global Synaptic Connectivity," Proceedings 12th International Workshop on cellular Nanoscale Networks and their Applications (CNNA), 2010.
[43] H. de Garis, C. Shuo, B. Goertzel, L. Rulting, "A World survey of artificial brain projects, Part I: Large-scale brain simulations", Neurocomputing 74, Issues 1-3, December 2010, pp 3-29.
[44] A. Parker et al, "Towards a Nanoscale Artificial Cortex", 2006 International Conference on Comuting in Nanotechnology, June 26, 2006, Las Vegas, USA.
[45] J. Joshi, J. Zhang, C. Wang, C. Hsu, A. Parker, "A Biomimetic Fabricated Carbon Nanotube Synapse with Variable Synaptic Strength," IEEE/NIH 5th Life Science and Applications Workshop, 2011.
[46] J. Patwardhan, C. Dwuyer, A. Lebeck, D. Sorin, "Circuit and System Architecture for DNA-Guided Self-Assembly of Nanoelectronics," Prcoeedings of Foundations of Nanoscience, ScienceTechnica, 2004.
[47] G. Snider, "From Synapses to Circuitry: Using Memristive Memory to Explore the Electronic Brain," IEEE Computer 44, 20, p21-28, 2001.
[48] E. Izhikevich and G. Edelman, Large-scale Model of Mammalian Thalmocortical systems, The Neurosciences Institute, San Diego, CA, 2007.
[49] G. Indiveri et al, "Neuromorphic Silicon Neuron Circuits," Frontiers in Neuroscience 5, 73, May 2011.
[50] A. Meltzoff, P. Kuhl, J. Movellan, T. Sejnowski, "Foundations for a new science of learning," Science 325: 284-288 (2009).
[51] E. Farquhar, C. Gordon: "A field programmable neural array," IEEE International Symposium on Circuits and Systems, May 2006.
[52] C. Mead, VLSI and Neural Systems, Addison-Wesley, 1989.
[53] S. Philipp, A. Grubl, K. Meier, J. Schemmel, "Interconnecting VLSI Spiking Neural Networks using Isochronous Connections," in Proc 99th International Work-Conference on Artificial Neural Networks, Springer LNCS 4507 pp 471-478, June 2007.
[54] E. Kandel, In Search of Memory, W.W. Norton, 2006
[55] J. Navaridas, M. Lujan, J. Miguel-Alonso, L. Plana, S. Furber, "Analysis of the Multicast Traffic Induced by Spiking Neural Networks on the SpiNNaker Neuromimetic System," Proceedings 23rd ACM Symposium on Parallelism in Algorithms and Architectures, San Jose, June, 2011.

Review Paper

Autonomy Rebuilt: Rethinking Traditional Ethics towards a Comprehensive Account of Autonomous Moral Agency

Jeffrey Benjamin White

Korea Advanced Institute of Science & Technology, Korea
**corresponding author*: jeffreywhitephd@gmx.com

Abstract

Autonomous agency is complex, bound up as it is with moral agency. And, moral agency is anything but clear. Confronted with many unanswered questions, researchers often operate under two distinct notions of autonomy, one associated with human and another with artificial agents. This lack of uniformity is theoretically unappealing, impedes progress on both forms of agency, and its constructive resolution is the focus of this paper. First, we review Kant's account of autonomous agency, and then turn to some contemporary analyses in which this robust understanding of autonomy is reduced to suit artificial applications. From this reduction, we review some contemporary approaches to understanding autonomy, thereby opening a way back to a comprehensive account of agency. And finally, we integrate the results of this discussion into a model of autonomous agency that can serve both as a platform for testing theories of moral decision and action, and as a framework for engineering and evaluating autonomous agents and agency.

Keywords: Autonomous agent, artificial intelligence, moral decision and action

1. Introduction

Autonomous agency is complex, bound up as it is with moral agency. Indeed, "a moral agent is necessarily an autonomous agent."(Smithers, 1997, page 95) And, moral agency is anything but clear, bound up as it is with things like freewill, responsibility, intention, conscience, personal identity and selfhood. Confronted with so many un-answered questions, researchers often operate under two distinct notions of autonomy, one associated with human and another with artificial agents. This lack of uniformity is theoretically unappealing, impedes progress on both forms of agency, and its constructive resolution is the focus of this paper.

Towards this end, Anthony Beavers (2012) suggests that the "hard problem" in morality lies in "rearranging" the landscape of traditional moral concepts so that solutions to problems in engineering artificial moral agents (AMAs) present themselves. The alternative on his account is the possible end of ethics, "ethical nihilism," with traditional moral concepts such as conscience and autonomy replaced only by the hollow objective determination of an agent's position in a chain of efficient causation. Meanwhile, Wendell Wallach (2010) suspects that the lack of progress in ethics is due to a preoccupation with isolable moral faculties, rather than "recognizing that moral acumen emerges from a host of cognitive mechanisms" and that "all of those considerations either merge into a composite feeling or conflict in ways that prompt the need for further attention and reflection," with the moral agent necessarily functioning as an "integrated being."(page 249) Wallach thereby calls for a "comprehensive" account of moral agency, one that can serve as "a platform for testing the accuracy or viability of theories regarding the manner in which humans arrive at satisfactory decisions and act in ways that minimize harms."(page 248)

It is my suspicion that Wallach's demands can be met through something like Beavers' means. The conceptual resources necessary for a comprehensive account of autonomous moral agency are available in traditional ethics, but have been hidden behind conventional interpretations and summarily established conceptions of human relative to artificial agency. The present paper attempts some moral landscaping to stop the erosion of ethics into transactional recordkeeping. First, it reviews some analyses in which autonomy is reduced, inviting ethical nihilism. Then, it clears the way to a comprehensive account of agency. Finally, it constructs such an account from traditional materials, resulting in a model of autonomous agency that is not only integrated, but integrative, and that can serve both as a platform for testing theories of moral decision and action, and as a framework for engineering and evaluating autonomous agents and agency.

2. Recognizing distinctions

Etymologically, the term "autonomous" is ancient Greek, with ''auto'' meaning self, and ''nomos'' meaning law. Originally, it applied to societies, cities, and states, which were considered autonomous when their members lived according to custom and convention specific to their common environment, thereby creating their own laws,

rather than having laws externally imposed. Autonomy thus means "self-governing."

Immanuel Kant developed this original notion of autonomy in terms of individual moral agency, with the model for the autonomous agent being "the political sovereign not subject to any outside authority, who has the power to enact law," and autonomous thereby meaning "self-sovereign."(Reath, 2006, page 122) From here, Kant specified that each is not only able to create and to act from laws of his own creation, to be "rational," but "to pass judgment upon himself and his own actions" from the ideal vantage point of a "kingdom of ends," an ideal arrived at through the exercise of moral duty, for every "man" "to make mankind in general his end,"(Kant, 1780, page 26) meaning that every rational agent should identify its own interests with the "kingdom of ends" in the mode of the moral equivalent of the political sovereign. Famously, this Kantian agent is guided by a single principle, the categorical imperative, one form of which commands an agent "Never to employ himself or others as a mean, but always as an end in himself,"(Kant, 1796, page 37) with "end in himself" meaning self-sovereign, and so qualitatively equivalent with the agent, itself.(see Kant, 1788, page 89)

Autonomy thus requires that the autonomous moral agent be free from selfish material desires, thereby embodying virtue worthy of "reverence," i.e. deserving of the respect and admiration cum emulation deserving of a beneficent king. "Autonomy is therefore the ground of the dignity of humanity, and also of every other intelligent nature whatsoever."(Kant, 1796, page 39)

However, it is exactly this degree of autonomy that is not afforded artificial agents in their very conception. Consider Ronald Arkin's 2009 text *Governing lethal behavior in autonomous robots*, with the obvious concern, how one "governs" "autonomous" agents, an equally obvious oxymoron. On this account, robot "autonomy" is limited self-direction toward goals of external origin within a human command hierarchy, i.e. serving as means to another's ends. After all, self-legislating warrior-robots acting to preserve dignity dignity, rather than blindly following orders to maim and murder, run counter to the intractable role that Arkin presumes violence playing in the press of history. We will have more to say about this presumption, and what it means to our conception of autonomy, in a moment. Regardless, on such account, AMAs are better understood as AAAs, artificial *amoral* agents, with robot autonomy rather rendered as *not*-autonomy, at all.

A similar reduction is effected by Michael Arbib (2005). On his assay, humans enjoy the dignity of self-determination with each "finding his or her own path in which work, play, personal relations, family, and so on can be chosen and balanced in a way that grows out of the subject's experience rather than being imposed by others," while for a robot "the sense is of a machine that has considerable control over its sensory inputs and the ability to choose actions based on an adaptive set of criteria rather than too rigidly predesigned a program."(Arbib, 2005, page 371) With artificial agents cast as objects of purely external determination rather than moral subjects, this is also a characterization of *not*-autonomous amoral agency. Finally, the rigid distinction between human and artificial moral agency is articulated by Tom Ziemke (2008) in terms of a Kantian inspired distinction between the "phenomenal" and "noumenal," with the first ascribed and the latter emerging via autopoietic self-organization, and with robots ultimately lacking the material constitution necessary to emerge as autonomous in the full sense, *not*-autonomous and so amoral by default.

In each preceding case, researchers propose two classes of autonomy so different that it is difficult to trace them to same concept at all. This divide can be smoothed over by rendering differences in autonomy by degree, however. Prima facie, there are three degrees of autonomous agency applicable to artificial agents. One, as a direct extension of human agency, only. This is a machine on auto-pilot, for example a landmine or a BMW on cruise control. Two, as an indirect extension of human agency. This is the conception most common to artificial agents, that they will act according to interred rules fed top-down, whether categorical principles or conditional guidelines. Arkin's military robots serve as good examples here; as human soldiers follow codes of warfare, so should their machines. The third degree specifies autonomy in the fully sovereign sense, representing both the promise of continued research in artificial intelligence – a fully autonomous AMA – and the promise of traditional moral education – autonomous human agents (AHAs) thriving in a just world of their own creation.

James Moor's is perhaps the most influential graduated analysis of autonomous moral agency.(Moor, 2006, 2007) At the lowest level, any agent or artifact the actions of which have ethical consequences qualifies as an "ethical impact agent." Moor offers the replacement of human jockeys with robotic jockeys in Qatar as an example here, whereby humans were freed from torturous servitude by machines unable to suffer similarly. One level higher, "implicit ethical agents" are morally significant by design. Moor's examples of such are spam-bots and airplane instruments that warn pilots of unsafe conditions, clearly degrees of currently realized ethical "agency," and still direct extensions of human agency. Moor's third type of ethical agent, the "explicit ethical agent," is able to identify morally salient information within specific contexts and to act according to appropriate principles. An indirect extension of human agency, Moor feels that this is the "paradigm case" of robot ethics, "philosophically interesting" and "practically important" while not too sophisticated to be realized. Finally, Moor's fourth type of ethical agent is the "fully ethical agent," by Moor's estimation not a level of agency likely to be realized in robots, representing self-sovereign agency characterized by three distinctly human characteristics - free will, consciousness, and intentionality - the engineering of which present serious problems.

Schermerhorn and Scheutz (2004) have also proposed a graduated classificatory schema. Theirs includes perceived

autonomy in the spirit of Ziemke's "phenomenal" autonomy. Their first degree of autonomy involves executing some function without direct human assistance.

Robotic jockeys qualify here, as would robotic vacuum sweepers. The second involves following human directives, without the need for step-by-step direction. Military missions would qualify as such directives, with this degree expressed by mission-capable agents. Schermerhorn and Scheutz's third level involves goal self-ascription and independent decision-making facilitated by self-reflective capacities over intentional states, corresponding to the fully ethical agent on Moor's hierarchy.

Finally, Schermerhorn and Scheutz point to a neglected aspect of autonomous agency, the perception and ascription of autonomy based on demonstrations of agency. For instance, in some situations, an autonomous agent will simply sweep the floor when put to that task, while in others it will stop sweeping to save the neighbor's cat from a burning barn, with this latter demonstration inviting an ascription of autonomy and the former, not.

The trouble here is that autonomy involves the context sensitive capacity to do the right things at the right times, and fully autonomous agents do not always appear that way, confounding any easy ascription on phenomenal bases.

Autonomy is "adjustable," and the demonstrated capacity to adjust the degree of autonomy that an agent expresses is essential to both autonomous agency and its ascription. Being a fully autonomous agent often involves ceding autonomy through "transfer-of-control," reflecting the fact that even fully autonomous agents pursue objectives of external derivation, e.g. as part of a team.(Pynadath et.al., 2002) Extending the context of team to include family, company, society, it becomes clear that most human action is externally determined, with original and on-going control over one's own "path in life" ceded well prior to birth and as a matter of course. In light of this fact, any distinction between human and artificial agent based on apparent source of guidance and goal may be misplaced. And this poses a real problem, not only for our ascription of autonomy to artificial agents, but for any conception of autonomous agency, at all.

For instance, consider that in the great team that is the military, both robots and humans are embedded within the same command hierarchy, in which "commanders must define the mission for the autonomous agent whether it be a human soldier or a robot."(Arkin, 2009, pages 37-38) Human and robot are equally embedded in this chain of command, with any failure to follow orders not revered as demonstrated moral virtue, but rather condemned as malfunction. Accordingly, to conceive of the human soldier as an autonomous agent in any non-contradictory sense requires a notion of autonomy inclusive of non-human killing machines, as well. Thus, we might reduce autonomous agency to "an embodied system designed to satisfy internal or external goals by its own actions while in continuous long term interaction with the environment in which it is situated."(Beer, 1995, page 173) But, this is just to say that an autonomous agent is simply a kind of efficient cause, that there is ultimately no distinction to be made between human and artificial agency, and that Beavers' fears of "ethical nihilism" have come true.

3. Rearranging the landscape

Interestingly, Kant also warned of the "quiet death" of morality, by the reduction of autonomy to "the physical order of nature."(Kant, 1780, page 7) But, before we review his defense of moral autonomy, it will pay to trace Ziemke's concept of this strong, "noumenal" autonomy typically reserved for humans to its origins in autopoiesis. "Autopoiesis," from the Greek meaning "self-producing," represents a rigid distinction between living and artificial agency according to which artificial agents are constitutionally incapable of autonomy, with an "allopoietic system like a robot deriving function from an external source," and the "primary function" of an autopoietic system "self-renewal through self-referential activity."(Amoroso, 2004, page 144) "Autopoiesis is the mechanism that imparts autonomy to the living,"(Luisi, 2003, page 52) with "the minimal form of autonomy" "a circular process of self-production where the cellular metabolism and the surface membrane it produces are the key terms."(Weber & Varela, 2002, page 115) So given, an autopoietic system is an organism that is both self-organizing, "one that continuously produces the components that specify it, while at the same time realizing it (the system) as a concrete unity in space and time, which makes the network of production of components possible,"(Varela, 1992, page 5) and far-from-equilibrium due to metabolic storage and "budgeting" of matter and energy in the development and maintenance of the "bodily fabric."(Boden, 1999, 2000)

We have already confronted some difficulties in distinctions based in the sources of goals, but there is much more to be said of "self-referential activity," a concept most important to the following section. According to an autopoietical account of agency, an organismThis bodily fabric emerges as a single, bound entity within "molecular space," with its properties (including semiological properties as signs and symbols are not abstract tokens but rather material tools) "structurally determined" by potential and actual chemical changes to the system.(Romesin, 2002) Co-emergent with the organism is the "niche," "the domain of interaction of the system with its surroundings, conditioning its possible ways of coupling with the environment,"(Rudrauf et.al., 2003, page 34) in terms of which it cognizes and acts in "selective coupling" with aspects of the environment, constituting the "operational closure" of the system., "the domain of interaction of the system with its surroundings, conditioning its possible ways of coupling with the environment."(Rudrauf et.al., 2003, page 34) This relationship dynamic between selective coupling and self-conditioning in a bubble of structurally determined significance constitutes the "operational closure" of the system, invitingleads to a view of cognition as "enacted," with the autopoietic system ultimately "creating its own world."(Luisi, page 58) So understood, an agent is a "self-producing coherence" bound to "maintain itself as a

distinct unity as long as its basic concatenation of processes is kept intact in the face of perturbations, and will disappear when confronted with perturbations that go beyond a certain viable range which depends on the specific system considered."(Varela, page 5)

Perturbations - "inputs" generally speaking - are responsible for two general classes of change, those within a "certain viable range" being "changes of state" through which the capacity of the system to self-organize, or adapt, is maintained, and "disintegrative changes" through which it is not.(Romesin) As "operational closure" extends from the molecular to cellular to organismic levels of organization, and upwards to social, cultural, and philosophical levels, an agent's niche can be understood as layers of increasingly conceptual order established in pro-active defense against disintegrative change, thus implying that "our minds are, literally, inseparable" not only from our bodies but from the environment as we experience it, thereby constituting a peculiar sort of "prison."(Rudrauf et.al., page 40).

As with the soldier cemented within a command structure, it is difficult to see how entrenchment within one's own "circular process of self-production" can ground autonomous agency in any non-contradictory sense. Effectively imprisoned in a semiological bubble of its own structural co-determination, this is as much as any artifact a portrait of *not*-autonomous agency. Comparatively, it seems that an agent with complete information about its embodied processes and origins, capable of specifying exact changes to its structure toward self-determined ends - swapping modules to suit particular purposes, as we might envision an artificial agent able to do - would enjoy greater autonomy than could any "living" thing. Thus, autopoiesis appears to be unnecessary for autonomous agency.

The case for an autopoietical foundation of autonomy is further weakened by the fact that the autopoietical distinction between living and non-living systems, and so the logic by which it "imparts autonomy to the living," is not very clear. For instance, Varela is reported to have not objected to the ascription of life to some synthetic molecular structures, Luisi's micelles, arguing that "our notion of life is heavily permeated by a religious bias (the notion of soul), which makes it difficult to freely use the word "life" for simple chemical systems," and that "Once one is liberated from these constraints, the term "life" may acquire a plainer and more usable meaning."(Luisi, page 58) However, in making this move, any necessary relationship between life and autonomy in any robust sense is severed.

In like spirit, one may argue for similar liberality regarding the term "autopoietic." Once liberated from constraints of cellular metabolism and surface membranes, autopoiesis can be fruitfully applied in the analysis of other systems including institutions and organizations (Goldspink and Kay, 2003, Hall and Dousala, 2010), legal systems (Vilaca, 2010) and social systems as a whole, (Leydesdorf, 1993) most famously through the work of Niklas Luhmann on whose account such systems are decidedly autonomous.(see Viskovatoff, 1999) Finally, with the autonomy of social systems, we are returned to the original, very plain and useful notion of autonomy with which this paper began.

In order to construct a comprehensive account of autonomy inclusive of both human and artificial agents while avoiding "ethical nihilism," however, we must review two further concepts from the autopoietical lexicon, "homeostasis" and "decoupling." A concept fruitfully developed by Antonio Damasio within the cognitive sciences, homeostasis (or better "homeodynamics") is the dynamic stability of a complex system achieved by balancing internal and external pressures through largely automated physical processes. On Damasio's account, as the cells of the body "gravitate" toward "fluid" states and away from "strained" "configurations of body state," they contribute to the "contents of feelings" as "both the positive and negative valence of feelings and their intensity are aligned with the overall ease or difficulty with which life events are proceeding."(Damasio, 2003, page 132) The positive association with objects that facilitate said stability transforms the world of objects into a space of value, such that "by the time we are old enough to write books, few if any objects in the world are emotionally neutral," with felt content rendered as "foundational images in the stream of mind" corresponding to "some structure of the body, in a particular state and set of circumstances."(pages 197 and 56)

Here, in the "gravitation" away from strained states, there is a basis for Wallach's "composite feeling" that is at the same time not limited to "living" systems. Consider, in this light, the molecule. The common representation of a molecule is that of a system sans strain, static and at rest. However, a more realistic image oscillates from strained configuration to strained configuration in dynamic equilibrium between forces internal and external. Now, a molecule doesn't "create its own world," but its presence does influence its environment, in special cases grounding the emergence of cellular and then organismic levels of organization, ultimately leading to evaluative content in the form of "the feeling of what happens," with even social systems emerging from "molecular space" by extension. This is not to say that "homeostasis" is the proper term for molecular dynamics. Rather, it is to say that everything in nature is a dynamic system, with homeostasis simply naming equilibrium seeking tendencies present in higher orders of organization. Following Alfred Kuhn (1974), we may suggest that all systems seek equilibrium in terms of their environments, and understand homeostasis as the general tendency for complex systems to compensate for forces of change while maintaining stability, integrity, and by extension even human dignity.

Indeed, it is this general tendency that ultimately grounds the emergence of autopoietical niches, themselves. Niches are spaces of cognition and action protective against forces of disintegrative change, fundamentally realized in Luisi's "living" micelles. A micelle insulates its interiority from potentially damaging external pressures, constituting a fundamental integrity "decoupled" from the environment, a proto-semiological bubble of self-production, effectively creating its own world within itself and of its own resources.

It is from this capacity to decouple from the environment, and not "life" however understood, that we can construct an account of strong autonomy equally inclusive of human and artificial agents. Following de Bruin and Kastner (2011), "decoupling" means "reducing direct effects of environmental stimulation and opening up possibilities for internally regulated behavior,"(page 10) thereby freeing an agent to act according to internal constraints rather than reflexively according to external triggers. These internal constraints extend throughout the range of agency, from chemical to symbolic, with capable agents creating their own purely conceptual worlds from their own cognitive resources. Decoupling thereby facilitates "hypothetical thought," a computationally demanding operation facilitated by formal constructs including counterfactuals and imperatives. "For example, hypothetical thought involves representing assumptions, and linguistic forms such as conditionals provide a medium for such representations."(Stanovish and Toplak, 2012, page 10)

Formal representations of hypotheticals further facilitate autonomy by representing situations potentially attainable through action and decoupled from an agent's structurally determined chemical-environmental entrenchment. This capacity to formally represent alternatives that guide action is "syntax autonomy." Syntax autonomy relies on "symbolic memory" through which agents gain "an element of dynamical incoherence with their environment (the strong sense of agency)."(Rocha, 1998, page 10) This formally mediated "incoherence" grounds the emergence of social and moral systems represented in theories of ethics and writs of history and law. Through these formal constructs, agents stipulate ends toward which they feel that actions should aim in a process "which involves the mutual orientation of agents in their respective cognitive domains to shared possibilities for future."(Beer, 2004, page 324) All told, this capacity to decouple from external pressures through symbolic mediation and to coordinate action to commonly beneficial ends over temporal limits far exceeding those of any constitutive agent is a powerful evolutionary force, known in traditional moral theory as "freewill."(see Juarrero, 2009)

4. Reinterpreting the tradition.

The preceding may seem to have strayed far from Kant's moral theory, when in fact we have merely plotted points for comparison in more recent discussions. For example, Kant anticipated the autopoietic distinction between life and artifact in terms of self-organization. In both, each part exists "by means of the other parts" as well as "for the sake of the others and the whole." However, in the natural organism "its parts are all organs reciprocally producing each other," so constituting "a whole by their own causality." Such an organized being is not a "mere machine, for that has merely moving power, but it possesses in itself a formative power of a self-propagating kind which it communicates to its materials though they have it not of themselves; it organizes them."(Kant, 1790, page 202) So understood, an organism is a "natural purpose" for Kant, "just the way we normally, prima facie and intuitively, view the living."(Weber and Varela, 2002, page 106)

However, also on Kant's account, far from autopoiesis imparting autonomy to the living, autonomy is hamstrung by self-productive requirements of the bodily fabric. "Life is the faculty a being has of acting according to laws of the faculty of desire."(Kant, 1788, footnote page 9) Meanwhile, autonomy, "autonomy of the will," or "freedom" as he variously calls it, "is a property of all rational beings," and to be free an agent must merely "regard itself as the author of its principles independent of foreign influences,"(Kant, 1785, pages 64-5) with such "foreign influences" including "the faculty of desire."

Accordingly, not only is life unnecessary for autonomy, it is a potential obstacle, calling into question the moral superiority presumed of human over artificial agents. Let's revisit the Kantian inspiration behind Ziemke's distinction between "noumenal" and "phenomenal" agency in this light. For Kant, when an agent conceives of itself as a "noumenon," he conceives of himself as a "thing in itself," "as pure intelligence in an existence not dependent on the condition of time," i.e. as if "immortal."(Kant, 1788 page 118) Here, we may understand "immortal" as free from the motivating necessities of embodiment, including the drives to maintain bodily integrity that so occupy the living agent, such that "he can contain a principle by which that causality acting according to laws of nature is determined, but which is itself free from all laws of nature."(page 118) So unfettered, an agent can focus on syntactic integrity, i.e. act in accord with the categorical imperative. This is why Kant equates autonomy of will with moral law.(see for example Kant, 1785, pages 62 and 66) "Autonomy of the will is that property of it by which it is a law to itself (independently of any property of the objects of volition)."(page 56) The difficulty for Ziemke's schema is that this independence is not necessarily observable, rendering, as we have already seen, any phenomenal ascription of autonomous agency suspect.

Digging deeper, there is in Kant a model of cognition and agency both accounting for these inner processes as well as giving us something to look for in ascribing autonomy. Excavating this model from traditional moral verbiage is difficult, however. To begin with, it is not enough for the Kantian moral agent to act freely toward just any ideal end, for instance out of purely scientific interest toward realizing the world as it is rather than as it appears, i.e. the "noumenal" rather than "phenomenal." Such agency is ultimately contingent on some "object of volition." Rather, on Kant's account, the ultimate promise of autonomous agency presents itself in the form of an "archetypal" world.(Kant, 1785, page 44) The archetypal world, variously referred to as the "kingdom of ends," the "summum bonum," the "supreme independent good," and even "God," is an ideal moral situation differing from the noumenal in that it is one with which an agent is "not in a merely contingent but in a universal and necessary connection," being the "destination" "assigned" by the moral law, "independent of animality," the "summum bonum" of the world.(Kant, 1788, page 165)

Kant's archetypal world appears to represent the mythical Christian "heaven," and such was in fact the model. However, Kant explicitly rejects the notion that recognition of any "God" is necessary for autonomy – and with this goes any requirement of a Christian "soul," for example.(Kant, 1788, page 133) Rather, Kant argues that an agent must merely hold three conceptions in order to be (potentially) autonomous: freedom (specifically, conceiving one's self as having the capacity to self-legislate, rather than serve bodily desires), immortality (conceiving one's self as if unbound by temporal constraints on the preceding), and God as the existence of a "supreme independent good."(page 137) "God" so understood is a destination, the archetypal end of action and "object of a will morally determined." Actions in accord with this ideal moral situation produce a deep moral pleasure subjectively realized as "harmony" with the extant realm consisting of all intelligent beings sharing in this ideal, with self-conception as "free" and "immortal" serving as limiting conditions on realizing this end.

Here, we are approaching an answer to Wallach's call for a comprehensive "platform for testing," noting that Kant also asks "What, then, is really pure morality, by which as a touchstone we must test the moral significance of every action?"(Kant, 1788, page 157) The key to answering this question lies in understanding how this "harmony" with the archetypal moral situation is possible for any "intelligent being," as it is this relationship that will finally bring Kant's model of moral cognition into the clear.

Intelligent being, synonymous with rational being, is the minimal condition for autonomy, the capacity to self-legislate. Autonomous action is determined by conceptions of law, rather than by "animality."(see Kant, 1788, pages 37 and 129) Thus, the Kantian portrait of agency is two-sided. One side is "immanent" through "transcendence," a "world of intelligence" and product of "intelligible being." The other is product of the immediate environment, animal attraction to "objects of volition" within the phenomenal world of sense.(page 108) These constitute two essential poles within the agent, one material and one ideal, as the Kantian agent "has two points of view from which he can regard himself, and recognize laws of the exercise of his faculties, and consequently of all his actions,"(Kant, 1785, page 70) and from which he may "pass judgment upon himself and his own actions."

As such, Kantian operational closure extends from the phenomenal world of appearances to the noumenal world of "things in themselves," understood as the archetypal world when one's own autonomous moral potential is fully realized. In conceiving of himself as free from material and temporal constraints, with an eye to the universal ideal situation the realization of which is his potential as an intelligent being, the agent "transfers himself in thought" "from the impulses of sensibility into an order of things wholly different from that of his desires in the field of sensibility," a situation in terms of which he does not imagine himself to be more comfortable, physically, but rather to have increased "intrinsic self-worth," a "better person" and "a member of the world of the understanding."(Kant, 1785, page 72) Accordingly, when we, as intelligent beings, "conceive ourselves as free, we transfer ourselves into the world of understanding as members of it and recognize the autonomy of the will with its consequence, morality."(page 70) "Moral pleasure" thus arises as an agent transcends embodied limitations and moves forward to the morally ideal "world of understanding" as his necessary and sufficient end of action.

Here, we find the ultimate bedrock of autonomous agency. The "fluid state" of one's self conceived as one's best possible self at once attuned to the best conceivable situation motivates the autonomous agent to realize that situation of its own freewill. The very possibility of morality arises in this realization, and Kant ties the survival of morality to its corresponding pleasure. Further, as the capacity to embody this condition is what gives autonomy to the autonomous, in this we have the terms to draw adequate distinction between degrees of agency, artificial or otherwise. Indeed, Kant writes that "No man is wholly destitute of moral feeling, for if he were totally unsusceptible of this sensation he would be morally dead."(Kant, 1780, page 30) Moreover, once this condition is realized, felt as a "good will," a moral agent is loathe to let it go, and regress into a relatively strained state. However, in order to understand why, we must review another concept from traditional moral theory, conscience.

5. Comprehending Autonomy

In Kant's words, conscience is "moral capacity" present as "an inward judge" "incorporated" into an autonomous agent's being from the position of the moral ideal, "as the subjective principle of a responsibility for one's deeds before God,"(Kant, 1780, page 41) i.e. from the perspective of the archetypal world. Conversely, to act contrary to the "dictates of conscience" produces a physical pain, "like grief, fear, and every other diseased condition," evidence of a proportional disharmony. So, even as moral pleasure reveals the possibility of morality, the self-disgust of inner discord reveals the possibility of immorality, providing a powerful motivation to morality, for "when a man dreads nothing more than to find himself, on self-examination, worthless and contemptible in his own eyes, then every good moral disposition can be grafted on it, because this is the best, nay, the only guard that can keep off from the mind the pressure of ignoble and corrupting motives."(Kant, 1788, page 163)

And, as conscience is of the fabric of rational agency, the moral duty to make "mankind in general his end" can be rewritten as "do nothing which by the nature of man might seduce him to that for which his conscience might hereafter torment him."(Kant, 1780, page 24) Thus conscience is the mechanism, or following Kant the "spring," of autonomous moral agency. Finally, so long as agency is conceived of as being bound by these two poles, good will and self-disgust, "ethical nihilism" is averted.

With this, we have in hand all of the necessary ingredients to answer Wallach's call for comprehensive ethics as "integrated being." The mechanism of this

integration is conscience. Conscience is nothing less than the mechanism of autonomous moral agency. It is the embodied capacity for the hypothetical comparison of one situation with others in terms with which the agent already cognizes and acts. It lays out possible ends of action as situations in which the agent should reach homeostasis, allowing for their comparison and relative evaluation, with the difference providing the motivation to move toward some rather than others. As the constitution of these hypotheticals proceeds from a limited sphere of individual experience, augmented by affective and effective mirroring as well as taught "top-down," the scope of conscience expands gradually over the course of operation. As terms increase, given sufficient resources, the agent may be able to balance greater numbers of dimensions, with associated dimensions bound together under single operators, simplifying the computational task. And, with the space of action mapped through this operation, conscience motivates the agent to seek situations with minimal strain between one's own and others' current and expected future situations, with the global minimum specified as the Kantian "summum bonum."

Some points of interest fall out of this portrait of autonomous moral cognition. For one thing, it naturalizes intension, understood as an internal, motivating and relatively evaluative felt strain, or tension, between conscientiously compared situations. It also naturalizes freewill, understood as embodied metabolic/energetic potential to construct and to act toward ends of one's own self-determination. These characterizations differ from those common to philosophy of mind, demanding accounts that cannot be fully developed here. However, they have been developed as aspects of the ACTWith model of moral cognition, and this model has been articulated in the contexts of model based reasoning and moral agency (White, 2010), psychopathy and moral psychology (White, 2012a), entropy and information ethics (White, 2012b) and autonomy in machine ethics (White, *in press*). For another thing, it is in terms of conscience that distinctions between degrees of autonomy can be consistently made. For example, Kant tells us directly that an agent would be merely "a marionette or automaton" without the tension between the sensible and the ideal made possible by conscience, with any sense of freedom a "mere delusion" deserving the name "only in a comparative sense, since, although the proximate determining causes are internal, yet the last and highest is found in a foreign land,"(Kant, 1788, page 102) i.e. not determined through conscience as a judge from the perspective of one's own projected moral perfection, but externally. Returning to the issues with which this paper began, this model of autonomous moral cognition answers Wallach's concerns about "integrated being," surprisingly enough by describing a being which integrates situations within the space of itself, in the process constituting the moral sentiment that is at once autonomy's signature. And, according to Beavers' proposal, it has been arrived at through some moral landscaping.

Through the preceding, it should be clear that most researchers in artificial agency go wrong in presuming that Kantian moral law must be pre-programmed as an explicit set of rules, when Kant takes great pains to show that the moral law co-emerges with the constitution of the rational agent. This constitution grounds autonomy, and with this fact the moral law emerges from a capacity to act regardless of material inclination, towards some universally good end, the guiding principle to which is formalized in Kant's categorical imperative. Accordingly, in response to Tonkins' (2009) "challenge to machine ethics," the real challenge in engineering fully autonomous AMAs lies in undoing prejudices stemming from misinterpretations of traditional ethical theory. The first step on this road to realize that these misunderstandings are only as temporary as are our personal commitments to them. Should autonomy be reduced to efficient causes, it is due only to our own lack of insight, our incapacity to free ourselves from our own embodied habits and conventions. So enslaved, so *not*-autonomous, it is no wonder that autonomy should forever remain a mystery.

References

[1] Amoroso, R.L. & Amoroso P.J. (2004) The Fundamental Limit and Origin of Complexity in Biological Systems: A New Model for the Origin of Life, in D.M. Dubois (ed.) *CP718, Computing Anticipatory Systems: CASYS03 - Sixth International Conference, Liege, Belgium,* August 11-16, 2003, New York: American Institute of Physics.

[2] Arbib, M.A. (2005). Beware the Passionate Robot, in Fellous, J.M., & Arbib, M.A. (ed.) *Who needs emotions? The brain meets the robot*. Oxford: Oxford University Press.

[3] Arkin, R. C. (2009). Governing lethal behavior in autonomous robots. Boca Raton: CRC Press.

[4] Beavers, A.F. (2012) Moral Machines and the Threat of Ethical Nihilism, appearing in Lin, P., Abney, K., & Bekey, G. A. *Robot ethics: The ethical and social implications of robotics*. Cambridge, Mass: MIT Press.

[5] Beer, R. (1995) A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72, 173-215.

[6] Beer, R. (2004). Autopoiesis and Cognition in the Game of Life. *Artificial Life*, 10, 3, 309-326.

[7] Boden, M. A. (1999). Is metabolism necessary? *British Journal for the Philosophy of Science*, 50, 2, 231.

[8] Boden, M.A. (2000). Autopoiesis and life. *Cognitive Science Quarterly*, 1, 117-145.

[9] de Bruin, L.C. & Kästner, L. (2011) Dynamic Embodied Cognition. *Phenomenology and the Cognitive Sciences.* Accessible at: http://www.springerlink.com/content/euxt674w7361j717/

[10] Damasio, A.R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Orlando, FL: Harcourt.

[11] Goldspink, C., & Kay, R. (2003). Organizations as self-organizing and sustaining systems: a complex and autopoietic systems perspective. *International Journal of General Systems*, 32, 5, 459-474.

[12] Hall, W.P., and Dousala, S. (2010) Autopoiesis and Knowledge in Self-Sustaining Organizational Systems. *4th International Multi-Conference On Society, Cybernetics And Informatics*. Orlando, Florida, USA.

[13] Juarrero, A. (2009) Top-Down Causation and Autonomy in Complex Systems. In Murphy, N.C., Ellis, G.F.R., & O'Connor, T. *Downward causation and the neurobiology of free will*. Berlin: Springer.

[14] Kant, I. (1780) *The Metaphysical Elements of Ethics*. trans. Abbott, T.K. Pennsylvania State University Electronic Classics Series (2005). http://www2.hn.psu.edu/faculty/jmanis/kant/metaphysical-ethics.pdf

[15] Kant, I. (1785) *Fundamental Principles of the Metaphysic of Morals*. trans. Abbott, T.K. Pennsylvania State University Electronic Classics Series(2010). http://www2.hn.psu.edu/faculty/jmanis/kant/Metaphysic-Morals.pdf

[16] Kant, I. (1788) *The Critique of Practical Reason*, trans. Abbott, T.K. Pennsylvania State University Electronic Classics Series (2010).

http://www2.hn.psu.edu/faculty/jmanis/kant/Critique-Practical-Reason.pdf
[17] Kant, I.,(1790/1914) Bernard, J.H., & Liberty Fund. *Kant's Critique of Judgement*. London: Macmillan. http://files.libertyfund.org/files/1217/Kant_0318_EBk_v6.0.pdf
[18] Kant, I. (1796). *The Metaphysics of ethics*. Edinburgh: T. & T. Clark. Available at: http://oll.libertyfund.org/title/1443 on 2012-01-24
[19] Kuhn, A. (1974). *The Logic of Social Systems.* San Francisco, CA: Jossey-Bass.
[20] Leydesdorff, L. (1997) Is society a self-organizing system? *Journal of Social and Evolutionary Systems*, 16,3, 331-349.
[21] Luisi, P. L. (2003). Autopoiesis: a review and a reappraisal. *Die Naturwissenschaften*, 90, 2, 49-59.
[22] Moor, J.H. (2006) The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21, 18–21.
[23] Moor, J.H. (2007) Taking the Intentional Stance Toward Robot Ethics. *APA Newsletter*, 6, 14-17.
[24] Reath, A. (2006). *Agency and autonomy in Kant's moral theory*. Oxford: Clarendon Press.
[25] Rocha, L.M. (1998). *Syntactic autonomy*. Los Alamos National Laboratory. Washington, D.C: United States. Dept. of Energy.
[26] Romesin, H. M. (2002). Autopoiesis, Structural Coupling and Cognition. *Cybernetics and Human Knowing*, 9, 5-34.
[27] Rose, S. (1998). *Lifelines: Biology beyond determinism*. Oxford: Oxford University Press.
[28] Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J. P., & Le, V. Q. M. (2003). From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. *Biological Research*, 36, 1, 27-65.
[29] Smithers, Tim (1997) Autonomy in Robots and Other Agents. *Brain and Cognition* 34: 88-106.
[30] Stanovich, K. & Toplak, M. (2012) Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*. http://www.springerlink.com/content/x461x01027625w35/
[31] Tonkens, R. A challenge for machine ethics. *Minds & Machines*, 19, 421–438, 2009.
[32] Wallach, W. (2010). Robot minds and human ethics: The need for a comprehensive model of moral decision making. *Ethics and Information Technology*, 12, 3, 243-250.
[33] White, J.B. (2010). Understanding and augmenting human morality: An introduction to the ACTWith model of conscience. *Studies in Computational Intelligence*, 314: 607-621.
[34] White, J.B. (2012a) An Information Processing Model of Psychopathy and Anti- Social Personality Disorders Integrating Neural and Psychological Accounts Towards the Assay of Social Implications of Psychopathic Agents. In *Psychology of Morality*. New York: Nova Publications.
[35] White, J.B. (2012b). Infosphere to Ethosphere: Moral Mediators in the Nonviolent Transformation of Self and World. *International Journal of Technoethics*, 2, 4, 53-70.
[36] White, J.B. (book chapter, in press) Manufacturing Morality: A general theory of moral agencygrounding computational implementations: the ACTWith model. In *Computational Intelligence*. New York: Nova Publications.
[37] Varela, F. (1992). Autopoiesis and a biology of intentionality. Appearing in *Proceedings of a workshop on Autopoiesis and Perception*, 4–14. ftp://ftp.eeng.dcu.ie/pub/alife/bmcm9401/varela.pdf
[38] Vilaca, G.V. (2010) From Hayek's Spontaneous Orders to Luhmann's Autopoietic Systems. *Studies in Emergent Order*, 3, 50-81.
[39] Viskovatoff, A. (1999). Foundations of Niklas Luhmann's Theory of Social Systems. *Philosophy of the Social Sciences*, 29, 4, 481-516.
[40] Weber, A., & Varela, F.J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and Cognitive Sciences, 1*, 2, 97-125.
[41] Ziemke, T. (2008). On the role of emotion in biological and robotic autonomy. *BioSystems*, 91, 401-408.

Human-Centred Multivariate Complexity Analysis

David Looney, Mosabber U. Ahmed, and Danilo P. Mandic*

Imperial College London, U.K.
**corresponding author*: d.mandic@imperial.ac.uk

Abstract

Signatures of changes in dynamical system complexity, in the response of a living system to their environment, are reflecte in both the within- and cross-channel correlations in observed physiological variables. Only a joint analysis of these heterogeneous variables yields full insight into the underlying system dynamics. We illuminate the abilities of the multivariate multiscale entropy (MMSE) method to model structural richness, and illustrate its usefulness in human centred applications - complexity changes due to constraints (cognitive load, stress).

1. Introduction

Physiological responses of a living organism are a widely-studied form of complex system [1]. The analysis is not trivial and is conducted from heterogeneous physiological variables, from spiking neuronal activity to respiratory waveforms, with different dynamics, degrees of coupling and causality, and at multiple temporal and spatial scales [2].

The 'complexity-loss' theory states that the severity of constraints on a living system, caused by e.g. illness or ageing, is manifested by changes in the complexity of its responses - a natural measure of structural richness [3]. Standard entropy measures, such as Shannon entropy, Kolmogorov-Sinai entropy or approximate entropy, assess signal regularity (cf. randomness) but not true system complexity - represented by coupled dynamics at different scales. The multiscale entropy (MSE) method evaluates univariate sample entropy across multiple temporal scales revealing long range correlations - a key property of complex systems [4], [5] - and has been applied to estimate the complexity changes in physiological time series for numerous applications (congestive heart failure, Alzheimers disease and postural sway dynamics [6], [7]).

The above methods only cater for single-channel data and therefore fail to account for dynamical relationships that exist *between* the physiological variables. This limits their potential in e.g. medical applications - a medic would routinely examine brain and heart responses as well as eye and muscle activity. The recent multivariate multiscale entropy (MMSE) method was developed *specifically* to cater for both the within- and cross-channel dependencies for any number of data channels [6], [7], revealing coupled dynamics not observable using standard single-channel estimates (Gestalt).

We here revisit MMSE and illuminate its use in both classic and novel human-centred applications, focusing on identifying complexity signatures from physiological recordings caused by increased cognitive load and stress. The approach is validated both for homogeneous and for heterogeneous multivariate physiological variables.

2. Multivariate Multiscale Entropy

Multivariate multiscale entropy (MMSE) estimation (Matlab code available from [8]) is performed by two steps [6], [7]:

1) The different temporal scales are define by coarse-graining (moving average) the p-variate time series $\{x_{k,i}\}_{i=1}^{N}, k = 1, 2, \ldots, p$, with N samples in each variate. For a scale factor ϵ, the corresponding coarse coarse-grained time series: $y_{k,j}^{\epsilon} = \frac{1}{\epsilon}\sum_{i=(j-1)\epsilon+1}^{j\epsilon} x_{k,i}$, where $1 \leq j \leq \frac{N}{\epsilon}$ and $k = 1, \ldots, p$.
2) The multivariate sample entropy, $MSampEn$, is evaluated for each intrinsic scale within the multivariate $y_{k,j}^{\epsilon}$, and is plotted as a function of the scale factor ϵ.

2.1. MSampEn Calculation

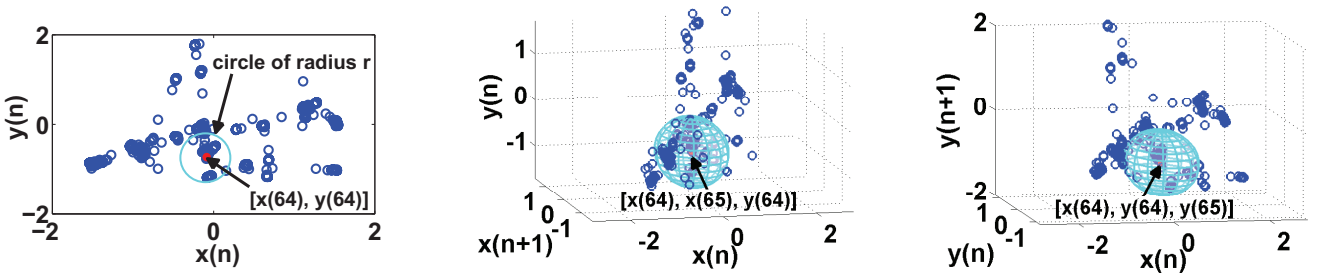
For a p-variate time series, $\{x_{k,i}\}_{i=1}^{N}, k = 1, 2, \ldots, p$ of length N, the calculation of $MSampEn$ is described in Algorithm 1 [6][7], where the multivariate embedded vectors are constructed as:

$$X_m(i) = [x_{1,i}, x_{1,i+\tau_1}, \ldots, x_{1,i+(m_1-1)\tau_1}, x_{2,i}, x_{2,i+\tau_2}, \ldots, x_{2,i+(m_2-1)\tau_2}, \ldots, x_{p,i}, x_{p,i+\tau_p}, \ldots, x_{p,i+(m_p-1)\tau_p}],$$

and $\mathbf{M} = [m_1, m_2, \ldots, m_p] \in \mathbb{R}^p$ is the embedding vector, $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_p]$ the time lag vector, and the composite delay vector $X_m(i) \in \mathbb{R}^m$ $(m = \sum_{k=1}^{p} m_k)$.

2.2. Geometric Interpretation of MSampEn

Underpinning the multivariate sample entropy method is the estimation of the conditional probability that two similar sequences will remain similar when the next data point is included. This is achieved by calculating the average number of neighbouring delay vectors for a given tolerance level (r)



(a) Scatter plot of [x(n), y(n)] (b) Scatter plot of [x(n), x(n+1), y(n)] (c) Scatter plot of [x(n), y(n), y(n+1)]

Figure 1. Geometry behind $MSampEn$. Scatter plots for a 2-variate gaze signal (see Fig. 3-A) for ($m = 2$) and the two 3-variate subspaces for ($m = 3$).

Algorithm 1: Multivariate sample entropy ($MSampEn$)

1) Form $(N-\delta)$ composite delay vectors $X_m(i) \in \mathbb{R}^m$, where $i = 1, 2, \ldots, N-\delta$ and $\delta = \max\{\mathbf{M}\} \times \max\{\boldsymbol{\tau}\}$ and defin the distance between any two vectors $X_m(i)$ and $X_m(j)$ as the maximum norm;
2) For a given composite delay vector $X_m(i)$ and a threshold r, count the number of instances P_i for which $d[X_m(i), X_m(j)] \leq r$, $j \neq i$, then calculate the frequency of occurrence, $B_i^m(r) = \frac{1}{N-\delta-1} P_i$, and defin $B^m(r) = \frac{1}{N-\delta} \sum_{i=1}^{N-\delta} B_i^m(r)$;
3) Increase $m_k \rightarrow (m_k + 1)$ for a specifi variable k, keeping the dimension of the other variables unchanged. Thus, a total of $p \times (N-\delta)$ vectors $X_{m+1}(i)$ in $\mathbb{R}^{m+1}$ are obtained;
4) For a given $X_{m+1}(i)$, calculate the number of vectors Q_i, such that $d[X_{m+1}(i), X_{m+1}(j)] \leq r$, where $j \neq i$, then calculate the frequency of occurrence, $B_i^{m+1}(r) = \frac{1}{p(N-\delta)-1} Q_i$, and defin $B^{m+1}(r) = \frac{1}{p(N-\delta)} \sum_{i=1}^{p(N-\delta)} B_i^{m+1}(r)$;
5) Finally, for a tolerance level r, estimate $MSampEn$ as

$$MSampEn(\mathbf{M}, \boldsymbol{\tau}, r, N) = -ln\left[\frac{B^{m+1}(r)}{B^m(r)}\right]. \quad (1)$$

and repeating the process after increasing the embedding dimension, from (m) to $(m+1)$, a geometric interpretation of which is shown in Fig. 1. Fig. 1(a) shows the set of delay vectors for a 2-variate gaze signal ([x(n), y(n)] with $\boldsymbol{\tau} = [1, 1]$ & $\mathbf{M} = [1, 1]$) and illustrates the neighbours[1] for the point [x(64), y(64)]. Upon increasing the embedding dimension from ($m = 2$) to ($m = 3$), we have two different subspaces spanning: (i) the vectors [x(n), x(n+1), y(n)] (Fig. 1(b)) and (ii) the vectors [x(n), y(n), y(n+1)] (Fig. 1(c)). The $MSampEn$ algorithm accounts fully for both within- and cross-channel correlations by examining the composite of all such subspaces.

[1]Neighbouring vectors at a point in an m-dimensional space can be represented by the points enclosed by an m-sphere or an m-cube, for the Euclidean and maximum norm respectively.

2.3. *Interpretation of the MMSE curves*

The complexity of multi-channel data is assessed from the MMSE plots ($MSampEn$ as a function of the scale factor). A multivariate time series is more complex than another one if for the majority of the time scales the multivariate entropy values are higher than those of the other time series. A monotonic decrease of the multivariate entropy values with the scale factor indicates that the signal in hand only contains useful information at the smallest scale and has no structure, and is therefore not dynamically complex (white noise).

3. Simulation Results

The potential of the MMSE method can be conveniently illustrated on a 6-variate time series, where originally all the data channels were realizations of mutually independent white noise. We then gradually decreased the number of variates that represent white noise (from 6 to 5, 3, 1 and 0) and simultaneously increased the number of data channels of independent 1/f noise (from 0 to 1, 3, 5 and 6). The 1/f noise exhibits long range correlations and is therefore complex [4], [5]. Fig. 2 shows that as the number of variates representing 1/f noise increased, the $MSampEn$ (complexity) at higher scales also increased, and when all the six data channels contained 1/f noise, the complexity at larger scales was the highest. The more variables/channels had long range correlations the higher the overall complexity of the underlying system - a key feature for use in human centred scenarios.

Fig. 3 illustrates that, unlike standard MSE, multivariate MSE caters for cross-channel correlations - a crucial advantage of the algorithm. Indeed, the complexity of the correlated bivariate 1/f noise with maximum correlation

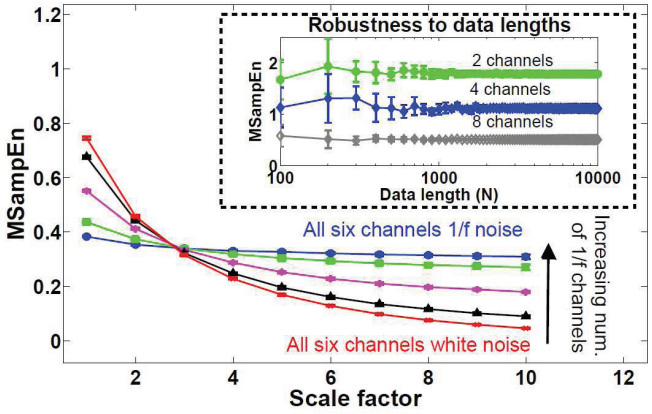


Figure 2. MMSE analysis for 6-channel data containing white and 1/f noise, each with 10,000 data points. The curves represent an average of 20 independent realizations and error bars standard deviation (SD). Figure insert in top right shows the robustness of $MSampEn$ estimates for white noise and illustrates that, the greater the number of channels, the more robust the $MSampEn$ estimate (errorbars become smaller).

coefficien ($cc = 1$) was the highest at large scales; the complexity decreased as the degree of correlation between the channels decreased and was lowest for the uncorrelated white noise.

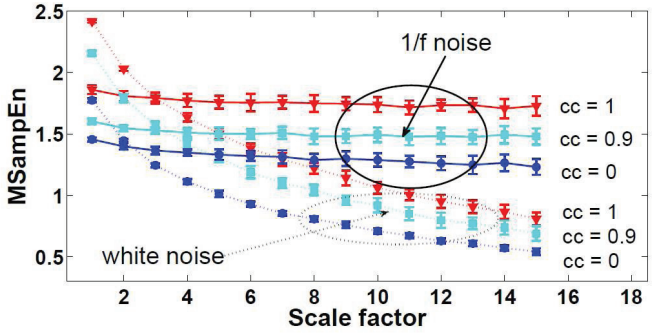


Figure 3. Multivariate multiscale entropy (MMSE) analysis for bivariate white and 1/f noise, each with 10,000 data points. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).

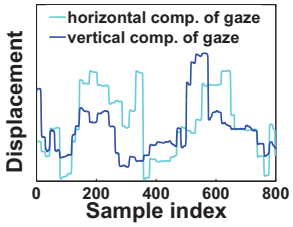
3.1. Analysis of Eye Gaze Dynamics

Psychologist Alfred L. Yarbus famously illustrated the impact of cognitive load on scanning eye patterns by presenting subjects with an image (see Fig. 4(a)) and recording gaze trajectories in response to different instructions [9]. To re-investigate this classic study from a completely novel perspective, we set out to examine whether cognitive load is reflecte in the complexity of the gaze dynamics. Seven healthy, naive subjects were asked to both examine the image in Fig. 4(a) freely and to complete six different instructions over 100 s trials (see [9] for more details), while bivariate (vertical and horizontal) eye gaze was recorded (a segment of which is shown in Fig. 4(c)).

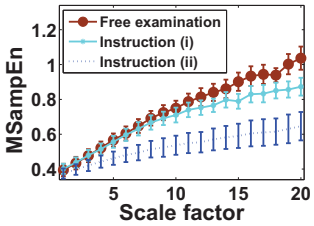
Fig. 4(d) shows the average gaze complexity over all subjects, for both constrained and free examination, and illustrates that the cognitive instructions can be uniquely identifie in the gaze complexity space. Compared to all instruction trials, the gaze complexity of free examination was the highest over high scale factors (>10), supporting the general 'complexity-loss' theory, that is, the less constrained the cognitive task the higher the complexity.

(a) Unexpected visitor

(b) Gaze intensity map



(c) Gaze signal



(d) Complexity analysis of gaze

Figure 4. Average MMSE analysis of the classic 'Yarbus experiment' illustrating that induced cognitive load reduces the gaze complexity. (a) The presented image. (b) Gaze intensity map for the instruction relating to the ages of the people. (c) Segment of raw gaze data, both horizontal and vertical components. (d) Average complexity results for 'free examination' and instructions (i) *estimate the material circumstances of the family* and (ii) *remember positions of people and objects in the room*, where the errorbars denote the standard error - std/sqrt(num. of trials).

3.2. Heart and Respiratory Function During Stress

Stress-induced illnesses are a major concern in modern mankind, the American Institute of Stress estimates that 75-90% of all visits to primary care physicians are for stress related problems. Stress is manifested by changes in several psycho-physiological modalities - a perfect match for the multivariate nature of the MMSE method.

Three naive, healthy subjects participated in a study (two parts each lasting 20 mins), in which respiration waveforms and electrocardiography (ECG) were recorded while the subject was seated comfortably and instructed not to talk or move unnecessarily. The baseline physiological response was established ('normal state') by engaging the subject in a relaxing task - watching a movie. Next, the subject was presented with a series of demanding mathematical logic questions and was instructed to respond via a keypad as quickly and accurately as possible ('stressed state'). An increased level of background noise and verbal interference from the experiment coordinator were used to increase the level of subject engagement.

Fig. 5 shows the average complexity results obtained for the bivariate data[2] [ECG, respiration] for both the 'normal' and 'stressed' states. The MMSE approach was clearly able to separate the two states in the complexity-space.

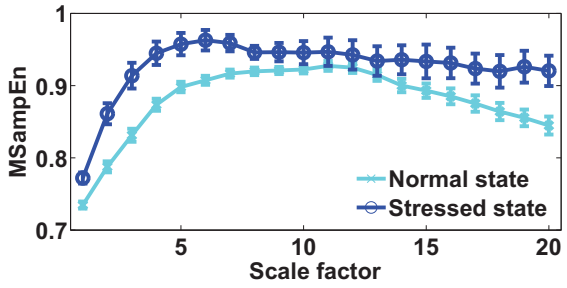


Figure 5. Average MMSE analysis for 'normal' and 'stressed' states, based on heart and respiratory functions. Error bars denote the standard error.

4. Conclusion

The recently introduced multivariate multiscale entropy (MMSE) method has been illuminated as an enabling tool for the complexity analysis of real-world multivariate data. It has been shown to model the dynamical couplings between physiological variables, giving an insight into the underlying system complexity, a feature not achievable using standard univariate measures. The advantages of MMSE have been exemplifie for detecting signatures caused by increased cognitive load and stress, highlighting its appeal in human-centred applications.

References

[1] K. Mainzer, *Thinking in Complexity: The Complex Dynamics of Matter, Mind, and Mankind*. Springer-Verlag, 1994.

[2] R. Gallagher and T. Appenzeller, "Beyond reductionism," *Science*, vol. 284, no. 5411, p. 79, 1999.

[3] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C. K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. Suppl 1, pp. 2466–2472, 2002.

[4] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, p. 068102, 2002.

[5] ——, "Multiscale entropy analysis of biological signals," *Phys. Rev. E*, vol. 71, no. 2, p. 021906, 2005.

[6] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Phys. Rev. E*, vol. 84, p. 061918, 2011.

[7] ——, "Multivariate multiscale entropy analysis," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 91–94, 2012.

[8] http://www.commsp.ee.ic.ac.uk/~mandic/research/Complexity.htm.

[9] A. L. Yarbus, *Eye Movements and Vision*. New York:Plenum, 1967.

[2]For each subject and sub-experiment, the data was divided into 100 s segments and at least 5 artifact-free segments were extracted and analysed. The ECG was bandpass filtere to occupy the frequency range (0.5 - 20) Hz and the respiration data was bandpass filtere to occupy the frequency range (0.05 - 3) Hz. All data was recorded at 1200 Hz and downsampled to 120 Hz.

2013 INNS Awards

By Leonid Perlovsky, Ph.D.
Chair of the Awards Committee of the INNS

The International Neural Network Society's Awards Program is established to recognize individuals who have made outstanding contributions in the field of Neural Networks. Up to three awards, at most one in each category, of $1000 each, are presented annually to senior, highly accomplished researchers for outstanding contributions made in the field of Neural Networks.

The Hebb, Helmholtz and Gabor Awards:
The Hebb Award - recognizes achievement in biological learning.
The Helmholtz Award - recognizes achievement in sensation/perception.
The Gabor Award - recognizes achievement in engineering/application.

Young Investigator Awards:
Up to two awards of $500 each are presented annually to individuals with no more than five years postdoctoral experience and who are under forty years of age, for significant contributions in the field of Neural Networks.

Nominations:

1. The Awards Committee should receive nominations of no more than two pages in length, specifying:
   -The award category (Hebb, Helmholtz, Gabor, or Young Investigator) for which the candidate is being nominated.
   -The reasons for which the nominee should be considered for the award.
   -A list of at least five of the nominee's important and published papers.
2. The curriculum vitae of both the nominee and the nominator must be included with the nomination, including the name, address, position/title, phone, fax, and e-mail address for both the nominee and nominator.
3. The nominator must be an INNS member in good standing. Nominees do not have to be INNS members. If an award recipient is not an INNS member, they shall receive a free one-year INNS membership.
4. Nominators may not nominate themselves or their family members.
5. Individuals may not receive the same INNS Award more than once

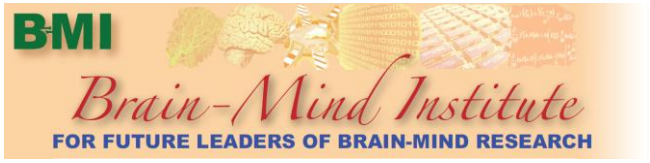
All nominations will be considered by the Awards Committee and selected ones forwarded to the INNS Board of Governors, along with the Committee's recommendations for award recipients. Voting shall be performed by the entire BoG.

The Awards Committee:
INNS Award Committee consists of the chair (Prof. Leonid Perlovsky) and two other members. All members must be INNS Governors in the year that they are appointed.

Please email the 2013 nominations along with their attachments directly to the chair of the Awards Committee at leonid@seas.harvard.edu, with a copy to the Secretary of the Society at jonathan@sit.kmutt.ac.th by June 1, 2012. Please use the following subject line in the email: INNS award nomination.

You may view this information at www.inns.org.

Call for Papers



Summer School
Mon. June 25 - Fri., August 3, 2012

International Conference on Brain-Mind (ICBM)
Sat. July 14, 2012 - Sun. July 15, 2012
Michigan State University, East Lansing, Michigan USA
http://www.brain-mind-institute.org/

Collectively, the human race seems ready to unveil one of its last mysteries — how its brain-mind works at computational depth. The research community needs a large number of leaders who have sufficient knowledge in at least six disciplines conjunctively — Biology, Neuroscience, Psychology, Computer Science, Electrical Engineering, and Mathematics (6 disciplines). The Brain-Mind Institute (BMI) provides an integrated 6-discipline academic and research infrastructure for future leaders of brain-mind research. The BMI is a new kind of institute, not limited by boundaries of disciplines, organizations, and geographic locations.

The *subjects of interest* include, but not limited to:

Genes: inheritance, evolution, species, environments evolution vs. development.

Cells: cell models, cell learning, cell signaling, tissues, morphogenesis, tissue.

Circuits: features, clustering, self-organization, brain areas, classification, regression.

Streams: pathways, intra-modal attention, vision, audition, touch, taste.

Brain ways: neural networks, brain-mind architecture, inter-modal, neural modulation (punishment/serotonin, reward/dopamine, novelty/Ach/NE, higher emotion).

Experiences/learning: training, learning, development, interaction, performance metrics.

Behaviors: actions, concept learning, abstraction, languages, decision, reasoning.

Societies/multi-agent: joint attention, swarm intelligence, group intelligence, laws.

Diseases: depression, ADD/ADHD, drug addiction, dyslexia, autism, schizophrenia, Alzheimer's disease, Parkinson's disease, vision loss, and hearing loss.

Applications: image analysis, computer vision, speech recognition, pattern recognition, robotics, artificial intelligence, instrumentation, and prosthetics.

Keynote talks include:

James L. McClelland, Stanford University
Stephen Grossberg, Boston University

INNS-WC2012
INNS-Winter Conference 2012
October 3-5, 2012, Bangkok, Thailand
http://inns.sit.kmutt.ac.th/wc2012/

The flagship conference of the International Neural Network Society (INNS) is the International Joint Conference on Neural Networks (IJCNN) that is jointly sponsored by INNS and IEEE Computational Intelligence Society. IJCNN traditionally features invited plenary talks by world-renowned speakers in the areas of neural network theory and applications, computational neuroscience, robotics, and distributed intelligence. In addition to regular technical sessions with oral and poster presentations, the conference program will include special sessions, competitions, tutorials and workshops on topics of current interest. Typically there are well over six hundred delegates in this annual event.

The board of governors of INNS decided in 2006 to establish a series of symposia or winter conferences devoted to new developments in neural networks. The first of the INNS Symposia Series was held in Auckland, New Zealand back on November 24-25, 2008 – http://www.aut.ac.nz/nnn08/. The theme was "Modeling the brain and the nervous system" and comprised of two symposia: 1) Development and Learning; and 2) Computational Neurogenetic Modelling. The second in the series was the INNS International Education Symposium on Neural Networks (INNS-IESNN) held in Lima, Peru on January 25-27, 2011 – http://eventos.spc.org.pe/inns-iesnn/index.html. The third Symposia Series will cover a much broader context of "Natural and Machine Intelligence".

INNS-WC general track: Trends in Natural and Machine Intelligence: Neural network theory & models; Computational neuroscience; Cognitive models; Brain-machine interface; Collective intelligence; etc.

INNS Symposium on Nature Inspired Creativity (SoNIC2012): Application of nature inspired computing in creative industries; Art and cognition; Generative art; Aesthetic evaluation; etc.

INNS Symposium on Vision and Image Processing (SoVIP2012): Low-level image processing; 3D sensing & object modeling; Tracking and surveillance; Human motion analysis; Robot intelligence; etc.

INNS Symposium on Data Analytics and Competitions (SoDAC2012): Business intelligence; Air quality and environmental issues; Social networks and analytics; Neuro-informatics; Data competitions; etc.

IJCNN2013

International Joint Conference on Neural Networks
August 4-9, 2013, Dallas, TX, USA
http://www.ijcnn2013.org/

The International Joint Conference on Neural Networks is the premier international conference in the area of neural networks. IJCNN 2013 is organized by the International Neural Network Society (INNS), and sponsored jointly by INNS and the IEEE Computational Intelligence Society - the two leading professional organizations for researchers working in neural networks.

IJCNN 2013 will be held at the Fairmont Hotel in Dallas, Texas. It will feature invited plenary talks by world-renowned speakers in the areas of neural network theory and applications, computational neuroscience, robotics, and distributed intelligence. In addition to regular technical sessions with oral and poster presentations, the conference program will include special sessions, competitions, tutorials and workshops on topics of current interest.

The beautiful city of Dallas is a hub of the high technology world. The Fairmont Hotel provides convenient access to all the attractions of Dallas, and is very close to several desirable destinations for attendees and their families.

Topics Covered

- Neural network theory & Models
- Brain-machine interfaces
- Evolving neural systems
- Learning neural networks
- Neurodynamics
- Neuroengineering
- Neural network applications
- Collective intelligence
- Self-aware systems
- Data Streams processing
- Agent-based systems
- Bioinformatics
- Computational neuroscience
- Cognitive models
- Embodied robotics
- Self-monitoring neural systems
- Neuroinformatics
- Neural hardware
- Pattern recognition
- Machine vision
- Hybrid systems
- Data mining
- Sensor networks
- Computational biology
- Artificial life

Organizing Committee

General Co-Chairs: Plamen Angelov and Daniel Levine
Program Chair: Péter Érdi
Program Co-Chairs: Marley Vellasco and Emilio del Moral Hernandez
Competitions Chair: Sven Crone
Tutorials Chair: Leonid Perlovsky
Special Sessions Chair: Radu-Emil Precup
Web Reviews Chair: Thomasz Cholewo
Panels Chair: Juyeng Weng
Publicity Chair: Bill Howell
Awards Chair: Arthur Kordon
Sponsors & Exhibits Chair: Jagannathan Sarangipani
Publications Chair: Bruno Apolloni
International Liaison: Carlo Morabito
European Liaison: Petya Koprinkova
Webmaster: Dan Alexandru

Plenary Speakers

- **Klaas Stephan**, University of Zurich, Swiss Federal Institute of Technology
- **Olaf Sporns**, Indiana University
- **Lydia Kavraki**, Rice University

Deadlines

- **Special Session, Tutorial Proposals:** December 15, 2012
- **Post-Conference Workshop Proposals:** December 15, 2012
- **Paper Submissions Deadline:** February 1, 2013
- **Camera-Ready Paper Submissions:** May 1, 2013
- **Early Registration:** June 15, 2013

INTERNATIONAL NEURAL NETWORK SOCIETY

IEEE Computational Intelligence Society