A Theory for Mentally Developing Robots

Juyang Weng

Embodied Intelligence Laboratory
Department of Computer Science & Engineering
Michigan State University
East Lansing, MI 48824, USA

Abstract

This paper introduces a theory about mentally developing robots. The limitation of the traditional agent model is raised and a new SASE agent is proposed, based on our SAIL developmental robot. We formulate the manual development paradigm and autonomous development paradigm. The performance of a developmental robot is then formulated as reaching the norm of a human age group. The framework of autonomously generating brain[1] *representation is investigated in mathematical terms. Some techniques of such a representation are provided based on our SAIL-2 developmental algorithm. We establish the conceptual limitation of symbolic representation and from the limitation we propose that no developmental robot can use a symbolic representation. Finally, the completeness of developmental robot is investigated conditioned on five factors.*

1 Introduction

In his pioneering paper published in 1950 titled "Computing Machinery and Intelligence" [1], Alan Turing envisioned a machine that can learn, which he called "child machine." He wrote:

> "Our hope is that there is so little mechanism in the child brain that something like it can be easily programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child."

However, there was a severe lack of computer controlled machinery, during his time when the first electronic computer Colossus had just been finished. Turing suggested in that paper a disembodied abstract machine and proposed an "imitation game," now called the Turing Test, to test it. The Turing Test had greatly influenced the modern day AI research that followed [2].

Not until the 1980's had the importance of embodiment received sufficient recognition in the AI community. The behavior-based approach, popularized by Rodney Brooks [3] and others [4][5], put situated embodiment back to the AI stage as it deserves.

However, *the mind* and its autonomous development, did not receive sufficient attention in the artificial intelligence community, until the late 1990's when SAIL robot [6][7] and Darwin V robot [8] started experiments on autonomous cognitive development. A 2001 article in Science [9] summarized the pivotal role that mental development should play in machine intelligence.

Traditional research paradigms in machine learning have been fruitfully informed by models of human learning. However, existing behavior-based learning techniques typically applied to robot learning (e.g., [10]) differ fundamentally from human mental development. Such differences are still not widely understood. Further, there is a need for basic theoretic framework for the new developmental paradigm.

This article intends to take up these issues. It does not discuss how to design a developmental program. The theoretical framework presented here is hopefully beneficial to answering some widely concerned conceptual questions and probably useful for designing developmental programs.

We first introduce a new kind of agent, the SASE agent, for mental development. Next, we study paradigms for constructing man-made systems, manual and autonomous. Section 4 discusses a formulation for cognition and behavior. Section 5 brings up the central issue of representation, and establishes the inapplicability of symbolic representation to mental development. Section 6 is dedicated to the completeness issue of the developmental paradigm in light of natural intelligence. Section 7 provides concluding remarks.

[1]The term "brain" is used for a developmental robot, but we do not claim that the brain of a developmental robot is similar to a biological one.

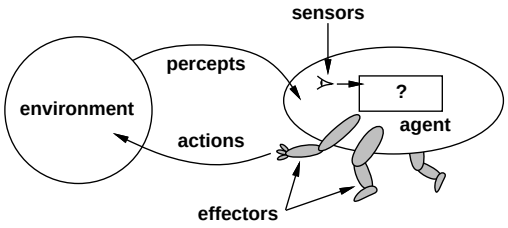

Figure 1. The abstract model of a traditional agent, which perceives the external environment and acts on it (adapted from Russell & Norvig [11]). The source of perception and the target of action do not include the agent brain representation.

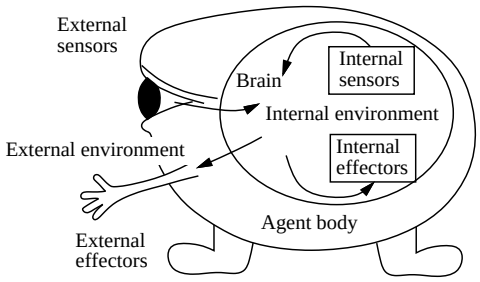

Figure 2. A self-aware self-effecting (SASE) agent. It interacts with not only the external environment but also its own internal (brain) environment: the representation of the brain itself.

2 SASE Agents

Defined in the standard AI literature (see, e.g., an excellent textbook by Russell & Norvig [11] and an excellent survey by Franklin[12]), an agent is something that senses and acts, whose abstract model is shown in Fig. 1. As shown, the environment E of an agent is the world outside the agent.

To be precise in our further discussion, we need some mathematical notation. A context of an agent is a stochastic process [13], denoted by $g(t)$. It consists of two parts $g(t) = (x(t), a(t))$, where $x(t)$ denotes the sensory vector at time t which collects all signals (values) sensed by the sensors of the agent at time t, $a(t)$ the effector vector consisting of all the signals sent to the effectors of the agent at time t. The context of the agent from the time t_1 (when the agent is turned on) up to a later time t_2 is a *realization* of the random process $\{g(t) \mid t_1 \leq t \leq t_2\}$. Similarly, we call $\{x(t) \mid t_1 \leq t \leq t_2\}$ a sensory context and $\{a(t) \mid t_1 \leq t \leq t_2\}$ an action context.

The set of all the possible contexts of an environment E is called the context domain $\mathcal{D}$. As indicated by Fig. 1, at each time t, the agent senses vector $x(t)$ from the environment using its sensors and it sends $a(t)$ as action to its effectors. Typically, at any time t the agent uses only a subset of the history represented in the context, since only a subset is mostly related to the current action.

The model in Fig. 1 is for an agent that perceives only the external environment and acts on the external environment. Such agents range from a simple thermostat to a complex space shuttle. This well accepted model played an important role in agent research and applications. Unfortunately, this model has a fundamental flaw: It does not sense its internal "brain" activities. In other words, its internal decision process is neither a target of its own cognition nor a subject for the agent to explain.

The human brain allows the thinker to sense what he is thinking about without performing an overt action. For example, visual attention is a self-aware self-effecting internal action (see, e.g., Kandel et al. [14], pp. 396 - 403). Motivated by neuroscience, it is proposed here that a highly intelligent being must be *self-aware and self-effecting* (SASE). Fig. 2 shows an illustration of a SASE agent. A formal definition of a SASE agent is as follows:

Definition 1 *A self-aware and self-effecting (SASE) agent has internal sensors and internal effectors. In addition to interacting with the external environment, it senses some of its internal representation as a part of its perceptual process and it generates actions for its internal effectors as a part of its action process.*

Using this new agent model, the sensory context $x(t)$ of a SASE agent must contain information about not only external environment E, but also internal representation R. Further, the action context $a(t)$ of a SASE agent must include internal effectors that act on R.

A traditional non-SASE agent does use internal representation R to make decision. However, this decision process and the internal representation R is not included in what is to be sensed, perceived, recognized, discriminated, understood and explained by the agent itself. Thus, a non-SASE agent is not able to understand what it is doing, or in other words, it is not self-aware. Further, the behaviors that it generates are for the external world only, not for the brain itself. Thus, it is not able to autonomously change its internal decision steps either. For example, it is not able to modify its value system based on its experience about what is good and what is bad.

It is important to note that not all the internal brain representations are sensed by the brain itself. For example, we cannot sense why we have interesting visual illusions [15].

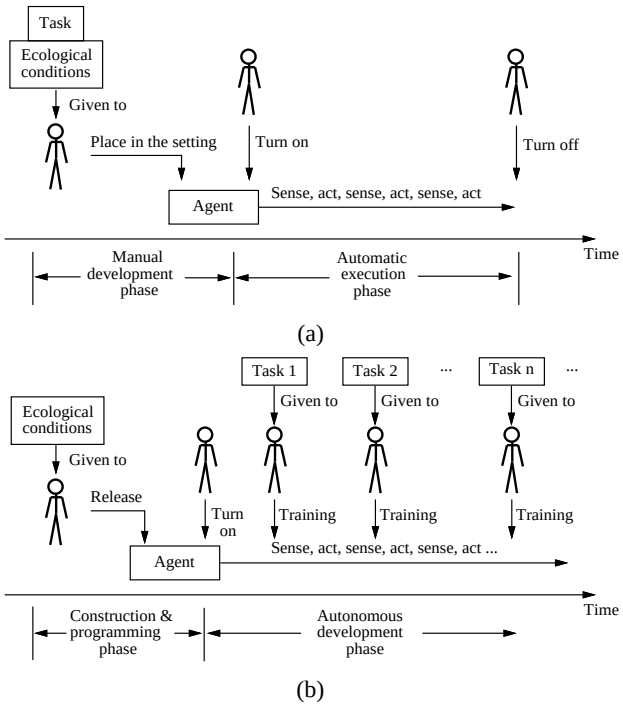

Figure 3. Manual development paradigm (a) and autonomous development (b) paradigm .

3 Machine Development Paradigms

The traditional paradigm for developing human engineered systems is manual. The distinctions between this traditional engineering paradigm and the new autonomous development paradigm are still not well understood. Here a mathematic formulation is proposed to delineate such distinctions.

An agent can perform one, multiple or an open number of tasks. The task here is not restricted by type, scope, or level. Therefore, a task can be a subtask of another. For example, making a turn at a corner or navigating around a building can both be a task.

Fig. 3 illustrates the traditional manual development paradigm and the autonomous development paradigm.

3.1 Manual development

The term "manual" refers to developing task-specific architecture, representation and skills through human hands. The manual paradigm has two phases, the manual development phase and the automatic execution phase. In the first phase, a human developer H is given a specific task T to be performed by the machine and a set of ecological conditions E_c about operational environment. The human developer first understands the task. Next, he designs a task-specific architecture and representation and then programs the agent A. In mathematical notation, we consider a human as a (time varying) function that maps the given task T and the set of ecological conditions E_c to agent A:

$$A = H(E_c, T). \tag{1}$$

In the second automatic execution phase, the machine is placed in the task-specific setting. It operates by sensing and acting. It may learn, using sensory data to change some of its internal parameters. However, it is the human who understands the task and programs its internal representation. The agent just runs the program.

3.2 Autonomous development

The autonomous development paradigm has two phases, first the construction and programming phase and second the autonomous development phase.

In the first phase, tasks that the agent will end up learning are unknown to the robot programmer. The programmer might speculate some possible tasks, but writing a task-specific representation is not possible without actually given a task. The ecological conditions under which the robot will operate, e.g., land-based or underseas, are provided to the human developer so that he can design the agent body appropriately. He writes a task-nonspecific program called *developmental program*, which controls the process of mental development. Thus the newborn agent $A(t)$ is a function of a set of ecological conditions only, but not the task:

$$A(0) = H(E_c), \tag{2}$$

where we added the time variable t to the time varying agent $A(t)$, assuming that the birth time is at $t = 0$.

After the robot is turned on at time $t = 0$, the robot is "born" and it starts to interact with the physical environment in real time by continuously sensing and acting. This phase is called autonomous development phase. Human teachers can affect the developing robot only as a part of the environment, through the robot's sensors and effectors. After the birth, the internal representation is not accessible to the human teacher.

Various learning modes are available to the teacher during autonomous development. He can use supervised learning by directly manipulating (compliant) robot effectors (see, e.g., [7]), like how a teacher holds the hand of a child while teaching him to write. He can use reinforcement learning by letting the robot try on its own while the teacher encourages or discourages certain actions by pressing the "good" or "bad" buttons in the right context (see, e.g., [16] [17]). The environment itself can also produce reward directly. For example, a "sweet" object and a "bitter" one (see, e.g., [18]). With multiple tasks in mind, the human

teacher figures out which learning mode is more suitable and efficient and he typically teaches one task at a time. Skills acquired early are used later by the robot to facilitate learning new tasks.

4 Robot Cognition in Continuous Context

How does a developmental robot "know" that it is given a task or it is performing a particular task? "Knowing" has a degree, and it depends on mental maturity. For example, at about 8 months of age, babies reach for an object in the last place they found it even when they have seen it moved to a new place. This is called A-not-B error [19]. Cognitive and behavioral capabilities gradually develop through experience.

Aristotle (448-380 BC) insisted that the mind is a "blank slate" at birth, a *tabula rasa*, which is, as we know now, not accurate according to the studies in developmental psychology [20] [21]. He is right, however, in recognizing that the experiences of the individual are of paramount importance and in his identifying the basic principle of association. Decartes's "rational approach" in the mid-1800's have been discarded by modern scientists, in favor of observational or empirical methods of studying the mind. How do we define and measure cognitive capabilities of our robots? Following the scientific tradition of careful quantification, clear definition and empirical observation, we propose the following framework for cognition by developmental robots.

First, cognition requires a discrimination among sensory inputs and a display of the discrimination through actions. Thus, we must address the concept of discriminative capability:

Definition 2 *Given a developmental agent at time t_1, suppose that the agent produces different action contexts a_1 and a_2, from two different contexts $C_1 = \{g(t) \mid t_1 \leq t \leq t_2\}$ and $C_2 = \{g(t) \mid t_1 \leq t \leq t_3\}$, respectively. If a_1 and a_2 are considered different by a social group (human or robot), conditioned on C_1 and C_2, then, we say that the agent* discriminates *two contexts C_1 and C_2 in the society. Otherwise, we say that the agent does not* discriminate *C_1 and C_2 in the society.*

The above definition allows variation of action context a from the same context C. In other words, even if different robots produce different actions in the same test, they are considered correct if the actions are considered socially equivalent. For example, no two humans have exactly the same voice but they can pronounce perceptually equivalent words.

It is not too difficult to discriminate just two particular contexts. We desire an agent to produce only equivalent actions from all the equivalent contexts. There is a special but very large class called the *unknown class* which includes all the contexts that the agent at this age is not expected to understand. Unlike a traditional classifier, we require a developmental robot to be able to deal with all possible contexts. That is, a development agent is supposed to produce a *correct action* even for contexts in D that it cannot deal with confidently. For example, if the context means "what is this?" the correct action for a baby robot can be "doing nothing" or, for a more mature robot, saying "I do not know" or anything else that is equivalent socially. Therefore, the extent of the context domain D is very important to the success of mental development. If the robot environment E is constrained significantly, the resulting context domain D might not be able to develop highly intelligent robots.

Definition 3 *Given a context domain $\mathcal{D}$ and a set of possible action contexts $\mathcal{A}$, a* norm *is a mapping N from D to $\mathcal{A}$, denoted by*

$$N : \mathcal{D} \mapsto \mathcal{A},$$

and it is defined by a social group. The agent mapping *of an agent at time t is also a mapping denoted by*

$$A(t) : \mathcal{D} \mapsto \mathcal{A}. \tag{3}$$

A test for an agent $A(t)$ is to let the agent experience multiple contexts. An evaluation of the performance is a measure that characterizes the agreement of the two mappings N and $A(t)$ through tests.

Since tasks are unknown during the robot programming time, during later mental development, a developmental program must generate an internal representation and some architecture "on the fly" for any task that the robot is learning, as shown in Fig. 3.

A mentally developing robot, or developmental robot for short, is an embodied, SASE agent that runs a developmental program following the autonomous developmental paradigm.

Different age groups of developmental robot have corresponding norms. If a developmental robot has reached the norm of a human group of age k, we can say that it has reached equivalent human mental age k.

5 Internal Representation

Autonomous generation of internal representation is central to mental development. The mapping $A(t)$ in Eq. (3) can be decomposed into three mappings: One is time varying *representation mapping* f_t, which generates internal representation $R(t)$, from context $g(t)$:

$$R(t) = f_t(g(t)). \tag{4}$$

The representation $R(t)$ is generated based on what we call developmental principles, from the experience indicated by $g(t)$. Some fine architecture is also represented in $R(t)$. $R(t)$ includes, for example, neurons (or equivalently, filters or feature detectors) and connections among the neurons. Well organized neurons form cortices which in turn form brain areas and regions (or equivalently, regression trees or mappings). In the SAIL-2 developmental program, the internal representation includes sensory mappings which use the staggered local PCA (principal component analysis) to approximate the function of early sensory cortices [22], and cognitive mappings that use incremental hierarchical discriminant regression (IHDR) [23] to approximate the function of later processing cortices.

The response computation function R_t computes the response of neurons in the representation $R(t)$ for the input context $g(t)$:

$$r(t) = R_t(g(t), R(t)).$$

The response corresponds to the firing signals of neurons in the cortices. If the representation changes, the dimension of response also does accordingly.

The time varying *action mapping* h_t maps context $g(t)$, response $r(t)$ of the representation $R(t)$ onto action context:

$$c_a(t) = h_t(r(t), R(t)). \quad (5)$$

The action mapping includes a hierarchy of value systems, which selects an action among multiple action candidates.

For simplicity, the above notation we used is in a batch fashion. Any practical developmental program must be incremental. Each sensory data $x(t)$ must be used to update the internal representation $R(t)$ which in turn is used for computing response $r(t)$ and action $a(t)$. As soon as it is done, $x(t)$ is discarded and the next sensory input $x(t+1)$ is received. Therefore, the entire context from birth to current time $g(t)$ is not available at time t and the developmental program cannot afford to store it either. The internal representation $R(t)$ is crucial to keep only necessary information about what has been learned so far. The incremental version of a developmental program can be represented by three mappings (we use the same mapping letters): Representation updating from current sensory input $x(t)$:

$$R'(t) = f_t(x(t), R(t)).$$

Response computation:

$$r(t) = R_t(x(t), R'(t)),$$

and action generation:

$$a(t) = h_t(r(t), R'(t)).$$

Fig. 4 shows how internal representation is incrementally updated and used by three mappings f_t, R_t and h_t. The

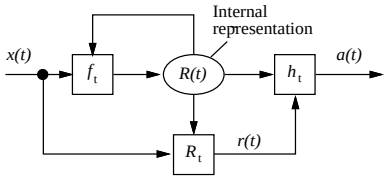

Figure 4. The three mappings inside a developmental agent: representation updating f_t, response computation R_t, and action generation h_t.

action $a(t) \in c_a(t)$ generated by the agent affects the world and, thus, affects the sensory input $x(t)$ in the next time instant. It is important to note that it is desirable that all effectors have a dedicated sensor so that the agent can sense what it is doing.

Next, we turn to the issue of the form of internal representation. Traditional AI systems use symbolic representation for internal representation and decision making. Is symbolic representation suited for a developmental robot? In the AI research, the issue of representation has not been sufficiently investigated, mainly due to the traditional manual development paradigm. There has been a confusion of concepts in representation, especially between reality and the observation made by the agents. To be precise, we first define some terms.

A *world concept* is a concept about objects in the external environment of the agent, which includes both the environment external to the robot and the physical body of the robot. The *mind concept*[2] is internal with respect to the nervous system (including the brain).

Definition 4 *A* world centered *representation is such that every item in the representation corresponds to a world concept. A* body centered *representation is such that every item in the representation corresponds to a mind concept.*

A mind concept is related to phenomena observable from the real world, but it does not necessarily reflect the reality correctly. It can be an illusion or totally false.

Definition 5 *A symbolic representation is about a concept in the world and, thus, it is* world centered. *It is in the form* $A = (v_1, v_2, ..., v_n)$ *where A (optional) is the name token of the object and* v_1, v_2, ..., v_n *are the* unique *set of attributes of the object with predefined symbolic meanings.*

For example, Apple $=$ (weight, color) is a symbolic representation of a class of objects called apple. Apple-1 $=$ (0.25g, red) is a symbolic representation of a concrete object called Apple-1. The set of attributes is unique in the sense that the object's weight is given by the unique entry

[2]The term "mind" is used for ease of understanding. We do not claim that it is similar to the human mind.

v_1. Of course, other attributes such as confidence of the weight can be used. A typical symbolic representation has the following characteristics:

1. Each component in the representation has a predefined meaning about the object in the external world.
2. Each attribute is represented by a unique variable in the representation.
3. The representation is unique for a single corresponding physical object in the external environment.

World centered symbolic representation has been widely used in symbolic knowledge representation, databases, expert systems, and traditional AI systems.

Another type of representation is motivated by the distributed representation in the biological brain:

Definition 6 *A distributed representation is not necessarily about any particular object in the environment. It is body centered, grown from the body's sensors and effectors. It is in a vector form $A = (v_1, v_2, ..., v_n)$, where A (optional) denotes the vector and v_i, $i = 1, 2, ..., n$ corresponds to either a sensory element (e.g., pixel or receptor) in the sensory input, a motor control terminal in the action output, or a function of them.*

For example, suppose that an image produced by a digital camera is denoted by a column vector I, whose dimension is equal to the number of pixels in the digital image. Then I is a distributed representation, and so is $f(I)$ where f is any function. A distributed representation of dimension n can represent the response of n neurons.

The *world centered* and *body centered* representations are the same only in the trivial case where the entire external world is the only single object for cognition. There is no need to recognize different objects in the world. A thermostat is an example. The complex world around it is nothing more than a temprature to it. Since cognition must include discrimination, cognition itself is not needed in such a trivial case. Otherwise, body centered representation is very different from a world centered representation. Some later (later in processing steps) body centered representations can have a more focused correspondence to a world concept in a mature developmental robot, but they will never be identical. For example, the representation generated by a view of a red apple is distributed over many cortical areas and, thus, is not the same as a human designed atomic, world centered symbolic representation.

A developmental program is designed after the robot body has been designed. Thus, the sensors and effectors of the robot are known, and so are their signal formats. Therefore, the sensors and effectors are two major sources of information for generating distributed representation.

Another source of information is the internal sensors and effectors which may grow or die according to the autonomously generated or deleted representation. Examples of internal effectors include attention effectors in a sensory cortex and rehearsal effectors in a premotor cortex. An internal attention effectors are used for turning on or turning off certain signal lines for, e.g., internal visual attention. Rehearsal effectors are useful for planning before an action is actually released to the motors. The internal sensors include those that sense internal effectors. In fact, all the conscious internal effectors should have corresponding internal sensors.

It seems that a developmental program should use a distributed representation, because the tasks are unknown at the robot programming time. It is natural that the representation in earlier processing is very much sensor centered and that in later processing is very much effector centered. Learned associations map perceptually very different sensory inputs to the same equivalent class of actions. This is because a developmental being is shaped by the environment to produce such a desired behavior.

On the other hand, an effector centered representation can correspond to a world object well. For example, when the eyes of a child sense (see) his father's portrait and his ears sense (hear) a question "who is he?" The internally primed action can be any of the following actions: saying "he is my father," "my dad," "my daddy," etc. In this example, the later action representation can correspond to a world object, "father," but it is still a (body centered) distributed representation. Further, since the generated actions are not unique given different sensory inputs of the same object, there is no place for the brain (human or robot) to arrive at a unique representation from a wide variety of sensory contexts that reflects the world that contains the same single object as well as others. For example, there is no way for the brain to arrive at a unique representation in the above "father" example. Therefore, a symbolic representation is not suited for a developmental program while a distributed representation is.

6 Completeness

The performance of a practical developmental robot is limited by the following five factors: (1) sensors, (2) effectors, (3) computational resource, (4) developmental program, and (5) how the robot is taught. For example, although the Dav robot body, shown in Fig. 5, has 43 degrees of freedom, it is far from a human body, even in just the degree of freedom alone. Further discussion of these practical issues is beyond the scope of this paper (the reader is referred to [6]). Here, we address the issue of fundamental theoretic limit of performance, assuming that the above five factors can improve continuously without a fixed bound.

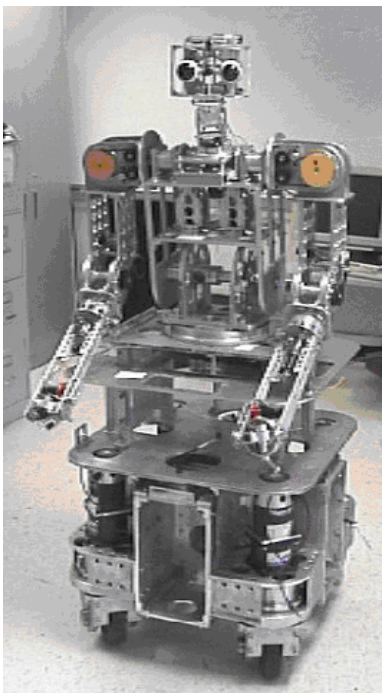

Figure 5. Dav developmental robot house-built at MSU.

Can a developmental robot understand any concept, say, "left," "right," "good," "bad," and a future concept "Internet-plus?" Any question of this nature does not entail a simple yes or no answer. Understanding has a degree. The higher the norm that an agent's mapping can reach on a subject, the more sophisticated its understanding is.

A natural question is: Can a developmental robot learn all possible new concepts? This is the completeness issue.

Definition 7 *A type of agent is conceptually complete if it can actually reach the human performance norm of any age group on any concept without human reprogramming.*

If a robot is not conceptually complete, it cannot learn all the concepts that a human can. We apply the restriction "without reprogramming" since human reprogramming allows the human to understand new concepts and then program into a machine, but the machine does not really understand it.

Why is this question important? The traditional AI is based on formal logic with a symbolic representation. Gödel [24] proved that certain questions cannot be answered correctly by any formal system. Turing [25] proved that there exists no program can decide correctly, in a finite number of steps, that any program P with input w will halt or run forever. Philosophers such as Lucas [26] have claimed that machines are inferior to humans because a human can "step outside" the limitation of logic. Mathematician Rogers Penrose [27][28] used Gödel theorem to support his view that any algorithmic procedure has a fundamental limit because mathematical "insight" that mathematicians use is not algorithmic. Since a developmental program cannot avoid being algorithmic in some sense, any developmental robot could run into a fundamental problem if Penrose was correct.

Can a machine go beyond formal logic? The fact that an algorithm is based on formal logic does not prevent it from "stepping outside" the formal logic. Humans are based on biology but that does not prevent humans from understanding biology from outside.

A major problem with the traditional manual development paradigm is that any machine from this paradigm is *world concept bounded*.

Definition 8 *A* programmed world concept *is a symbolic representation $A = (v_1, v_2, ..., v_n)$, where the meaning of some or all the attributes are defined by the human programmer.*

In the manual development paradigm, internal representation about the world is programmed in. Thus, such a representation gives only programmed world concepts, whether learning is used later or not. Typically, a programmed world concept A corresponds to the output from a human programmed procedure P that take sensory and effector signals as input and the attributes of A as output. The human programmer must know the meaning of A, otherwise, there is no way for him to program P.

Definition 9 *An agent is* conceptually bounded *if there exists a finite set of programmed world concepts $D = \{c_1, c_2, ..., c_n\}$, so that all the attributes of internal representation are functions of the concepts in D only.*

In other words, D includes all the dependent "variables" of internal representation. For example, if $D = \{\text{apple}, \text{banana}\}$. The the internal representation can constrain any function of "apple" and "banana," but it cannot create a new concept, say, "pear."

We have the following theorem:

Theorem 1 *Any agent developed by the manual development paradigm is* conceptually bounded.

Proof. In the manual development paradigm, the concepts are designed by the programmer based on the given task. During a limited time of programming, he can only design a finite concept set D. All the internal results generated by the robot during the automatic execution stage must be dependent on D only. Thus, any robot from the manual development paradigm is conceptually bounded. □

Obviously, a conceptually bounded robot is not conceptually complete since it cannot learn any new concept that is not included in the world concept set D. This result is stated as the following theorem:

Theorem 2 *Any agent developed by the manual development paradigm is not conceptually complete.* □

As Roger Penrose pointed out, humans can invent new mathematics. They do so based on new observations, and new concepts that are based on new observations. When they make new observations and discuss with colleagues, they are *mentally developing*. They form new concepts based on new phenomena and name them using new words or new word combinations that we have not used yet. They do use algorithmic rules, but those rules have new meanings when they are applied to new concepts.

Now, a question that is somewhat more difficult to answer is: "Is a developmental robot conceptually complete, if the capability of the five factors is not limited?" Note that we do not define a concept in terms a meaning understood by a developmental robot. The way a developmental robot treats a concept is by generating behaviors externally from the context that related to the concept. Thus, the concept here includes anything, correct and false, known by human now and those that will be known later.

Theorem 3 *A developmental robot is conceptually complete, if all required albeit finite capabilities of the five factors are satisfied.*

Proof. The proposed new paradigm for a developmental agent aims to realize a mapping of an embodied agent, from context space $\mathcal{D}$ to action space $\mathcal{A}$. To prove the theorem, we realize that a human being is a time varying (biological) mapping $A(t)$ as that in Eq. (3). An agent is conceptually complete if it can reach the norm of human performance of any age group. Suppose that human-level sensors, human-level effectors, human-level computational resource, human-level developmental program and human teaching environment are all given. If human agents can reach the performance about a given concept, a robot can too, provided that the five factors collectively give a set of sufficient conditions. In fact, they do, since they partition the space and time of an agent into five parts, nothing is left out: at the birth time, the robot is divided into four parts: sensors, effectors, computational resource and developmental program. After the birth time, the experience of the robot is included in the fifth factor: how the robot was taught. □.

Yes, it takes a lot of future work in terms of the five factors to greatly raise the norm of developmental robots. The most conceptual challenging factor is probably the developmental program. Other four factors are also extremely challenging. Nevertheless, a developmental robot, theoretically described here, is conceptually complete, compared with human, in the sense discussed above.

Therefore, a developmental robot is able to step outside the formal logic system. It is also able to step outside of any scientific subject. It is able to invent new mathematics, if it has reached the norm of the adult human age group. It can do all these because it can form new representation internally based on sensory inputs and effector experiences, just like what a human does in the same situation. The essence is "mentally developing" for new observations and concepts. This is not realizable by just adjusting parameters from a static human designed representation, nor by switching systems from a fixed number of human designed static formal systems, since they still correspond to a single fixed D.

Designing and evaluating tests for human children is a major task of a field called psychometrics. As Howard Gardner put in his book *Multiple Intelligences* [29], human intelligence is multiple, including linguistic, logical mathematical, musical, bodily kinesthetic, spatial, interpersonal and intrapersonal. We argued that developmental capabilities can serve as a unified metrics for machine intelligence [30]. It is expected that psychometric tests for human children, such as *Bayley Scales of Infant Development* [31] would be used for testing mentally developing robots.

7 Conclusions

This paper does not describe how to construct a developmental robot, nor how to write a developmental program. Instead, it addresses some basic theoretical issues that have not been solved by the traditional mental development paradigm and have not been established for the new autonomous development paradigm. Starting from a new SASE agent model, we propose that a mentally developing robot is not a function of tasks, but only of ecological conditions. This task independent property allows us to formulate mental development as incremental approximation for mapping from the space of contexts to the space of action contexts. This embodied, task-independent definition of the goal of mental development is consistent with that of biological mental development. We are careful not to introduce a value function since any such function has unavoidable fundamental limitations. The goal of life varies from person to person. It is hard for many people to accept that the goal of life is to spread genes [32].

In the past, the difference between symbolic representation and distributed representation used by neural network models and regression methods (such as SAIL [23]) was considered by many researchers as simply a matter of style or preference. We argued in this paper that in the new autonomous development paradigm, the symbolic representation is not suitable, giving a support for the use of biologically motivated, non-symbolic, non-logic, non-world semantic type of representation, which we called distributed representation here.

Finally, we discussed the completeness of developmental robots and we established that any agent resulted from the traditional manual development paradigm is concept

bounded but the model of developmental robots has no theoretical limit in reaching the human performance norm in the future. It is expected that future human teachers will play an important rule in teaching the future developmental robots, since the robot developmental learning will depend on the vast amount of human knowledge accumulated through many generations of human history. Of course, practical issues of the five factors are of great challenge to our future researchers.

Acknowledgments

The work was supported in part by NSF under grant number IIS-9815191, DARPA ETO under contract No. DAAN02-98-C-4025, DARPA ITO under grant No. DABT63-99-1-0014.

References

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

[2] E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought*, AAAI Press and The MIT Press, Menlo Park, CA and Cambridge, MA, aaai edition, 1995.

[3] R. A. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. 2, no. 1, pp. 14–23, March 1986.

[4] R. C. Arkin, *Behavior-Based Robotics*, The MIT Press, Chambridge, MA, 1998.

[5] L. Steels and R. Brooks, Eds., *The Artificial Life Route to Artificial Intelligence*, Lawrence Erlbaum, Hillsdale, NJ, 1995.

[6] J. Weng, "Learning in image analysis and beyond: Development," in *Visual Communication and Image Processing*, C. W. Chen and Y. Q. Zhang, Eds., pp. 431 – 487. Marcel Dekker, New York, NY, 1998, Also MSU CPS Tech. Report CPS-96-60, 1996.

[7] J. Weng, W. S. Hwang, Y. Zhang, and C. Evans, "Developmental robots: Theorey, method and experimental results," in *Proc. 2nd International Conference on Humanoid Robots*, Tokyo, Japan, Oct. 8-9 1999, pp. 57–64.

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

[9] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, "Autonomous mental development by robots and animals," *Science*, vol. 291, pp. 599–600, Jan. 26 2001.

[10] H. Hexmoor, L. Meeden, and R. R. Murphy, "Is robot learning a new subfield? the robolearn-96 workshop," *AI Magazine*, pp. 149–152, Winter 1997.

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

[12] S. Franklin and A. Graesser, "Is it an agent, or just a program?: A taxonomy for autonomous agents," in *Proc. the 3rd Int'l Workshop on Agent Theories, Architectures, and Languages*, New York, 1996, Springer-Verlag.

[13] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, 2nd edition, 1976.

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

[15] D. M. Eagleman, "Visual illusions and neurobiology," *Nature Reviews: Neuroscience*, vol. 2, no. 2, pp. 920–926, Dec. 2001.

[16] J. Weng, W. S. Hwang, Y. Zhang, C. Yang, and R. Smith, "Developmental humanoids: Humanoids that develop skills automatically," in *Proc. IEEE International Conference on Humanoid Robots*, Cambridge, MA, Sept. 7-8 2000.

[17] Y. Zhang and J. Weng, "Grounded auditory development by a developmental robot," in *Proc. INNS-IEEE International Joint Conference on Neural Networks*, Washington, DC, July 14-19 2001, pp. 1059–1064.

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

[19] J. Piaget, *The construction of reality in the child*, Basic Books, New York, NY, 1954.

[20] J. H. Flavell, P. H. Miller, and S. A. Miller, *Cognitive Development*, Prentice Hall, New Jersey, third edition, 1993.

[21] M. Domjan, *The Principles of Learning and Behavior*, Brooks/Cole, Belmont, CA, fourth edition, 1998.

[22] N. Zhang, J. Weng, and X. Huang, "Progress in outdoor navigation by the SAIL developmental robot," in *Proc. SPIE Int'l Symposium on Intelligent Systems and Advanced Manufacturing*, Newton, MA, Oct. 28 - Nov. 2 2001, vol. 4573.

[23] W. S. Hwang and J. Weng, "Hierarchical discriminant regression," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1277–1293, 11 2000.

[24] K. Gödel, "Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I," *Monatshefte für Mathematik und Physik*, vol. 38, pp. 173–198, 1931.

[25] A. M. Turing, "On computable numbers with an application to the Entscheidungsproblem," *Proc. London Math. Soc., 2nd series*, vol. 42, pp. 230–265, 1936, A correction, *ibid.*, 43, pp. 544-546.

[26] J. R. Lucas, "Minds, machines and Gödel," *Philosophy*, vol. 36, pp. 112–127, 1961.

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

[28] R. Penrose, *Shades of the Mind: A Search for the Missing Science of Consciousness*, Oxford University Press, Oxford, 1994.

[29] H. Gardner, *Multiple intelligences: The theory in practice*, Basic Books, New York, NY, 1993.

[30] J. Weng, "Autonomous mental development and performance metrics for intelligent systems," in *Proc. Measuring the Performance and Intelligence of Systems*, Gaithersburg, MD, Aug. 14-16 2000, pp. 349–358.

[31] N. Bayley, *Bayley Scales of Infant Development*, Psychological Corp., San Antonio, TX, 2nd edition, 1993.

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