New Ideas in Psychology 29 (2011) 217–227

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Computational modeling/cognitive robotics complements functional modeling/experimental psychology

Sidney D'Mello*, Stan Franklin

Department of Computer Science, 209 Dunn Hall, University of Memphis, Memphis, TN 38152-3230, United States

Keywords:
Cognitive robotics
Learning
Developmental robotics
Cognitive architectures
LIDA

ABSTRACT

This position paper explores the possible contributions to the science of psychology from insights obtained by building and experimenting with cognitive robots. First, the functional modeling characteristic of experimental psychology is discussed. Second, the computational modeling required for cognitive robotics is described, and possible experiments with them are illustrated. Next, we argue that cognitive developmental robots, robots that "live" through a development phase where they learn about their environments in several different modes, can provide additional benefits to the science of psychology. Finally, the reciprocal interactions between computational modeling/cognitive robotics and functional modeling/experimental psychology are explored. We conclude that each can contribute significantly to the other.

Experimental science progresses via a theorize → predict → experiment → theorize cycle during which a theory is built, predictions made from the theory are tested by experimentation, and the theory is revised in light of empirical findings, tested again, etc. (Beveridge, 1957; Losee, 1980; Salmon, 1990). All scientific theories are, to some extent, both functional and mechanistic in nature. A *functional* theory describes *what* can be expected to occur in a given situation. A *mechanistic* theory speaks to the *how* of the occurrence, the mechanism that brings it about.

Psychological theories are typically functional in nature (Angell, 1907; Block, 1980; Meissner, 1966). Psychologists build functional models that are intended to both explain psychological processes and predict their functionality, that is, what can be expected to happen under various conditions. Although

* Corresponding author. Tel.: +1 901 678 2146; fax: +1 901 678 2579.
E-mail address: sdmello@memphis.edu (S. D'Mello).

doi:10.1016/j.newideapsych.2009.07.003

these functional models are useful, even essential to understand human cognition, they do not reliably yield insight into the mechanisms underlying the cognitive processes. Additionally, they typically model only small pieces of cognition (e.g. memory and attention; Logan, 2002).

In contrast, the control system of any robot, by its very nature, must be fully integrated. That is, it must choose its actions based on incoming exogenous or endogenous stimuli utilizing *all* needed internal processes. Also, the control system of a robot must act through its underlying mechanisms (i.e. sensors and effectors). Almost by definition, the architecture of a cognitive robot, a robot that employs a cognitive architecture to select its next action, is derived from integrated models of the cognition of humans and/or other animals. Its control system is designed using that integrated cognitive architecture and is structurally coupled to its underlying mechanisms. Because the control system of such a robot cannot be restricted to small pieces of cognition, it would have to be sufficiently broad so as to encompass basic cognitive processes such as perception, episodic memory, selective attention, action selection, and action execution. Higher-level cognitive processes such as deliberation, volition, problem solving, developmental learning, and metacognition may be modeled as well. Additionally, it may be necessary to establish computational frameworks for feelings and emotions to serve as motivators and learning facilitators.

Like other experimental scientists, a roboticist may work through a theorize → experiment → theorize cycle. A robot is designed and built, but experimentation shows that it does not perform as desired. Therefore, its control system and underlying mechanisms are redesigned and rebuilt. More experimentation takes place resulting in more redesigning, etc.

The two theorize → experiment → theorize cycles (of experimental psychology and cognitive robotics) can be amalgamated by means of a cognitive robot that is able to participate in or replicate a psychological experiment. The cognitive architecture of the robot would functionally model the psychological process being experimented with on humans or animals. The computational architecture is essentially the same model acting through the underlying mechanisms. The computational architecture yields insight into the mechanisms underlying the process. The human or animal experiments together with the cognitive robot experiment serve to test both the functional model and the computational model. Both the high-level functional model and the underlying computational model can then be brought more in line with the results of these experiments. After alterations to the robot suggested by the new version of the architecture are made, new psychological experiments can be designed and carried out to test the current version. The amalgamated cycle continues.

The overall goal of this paper is to explore the possible contributions to the science of psychology from insights obtained by building and experimenting with cognitive robots. We begin by discussing several of the benefits to psychology afforded by computational models of human cognition. We then describe several shortcomings associated with computational modeling of human cognition that can be alleviated by the use of cognitive robotics. The paper then evaluates possible design paradigms for the control structure of a cognitive robot and argues for basing the control mechanism on computational models that are consistent with known psychological evidence. We then describe a number of "new" AI (artificial intelligence) techniques to simulate several of the cognitive processes required for the development of a cognitive robot that may be used in experiments to advance our understanding of human cognition. The paper then argues that cognitive robots that "live" through a development phase, during which they learn about their environments in several different modes can provide additional benefits to the science of psychology. Next we briefly discuss the feasibility of cognitive robotics and highlight several robotic simulation environments that may be useful for rapid development and experimentation. Finally, we explore the reciprocal interactions between computational modeling/cognitive robotics and functional modeling/experimental psychology.

1. Cognitive robotics as an extension of computational modeling

The fundamental goal of cognitive science is to obtain a better understanding of human cognition. Towards this end, psychologists and cognitive scientists have proposed theories and hypotheses, formulated research questions, designed and conducted controlled experiments to evaluate such hypotheses while systematically addressing confounds and alternatives, and have subsequently accepted, rejected, or revised their theories.

Computational models have long been a major, possibly indispensable tool in science (Barto, 1991; Bower, 1992; Maynard Smith, 1974). In cognitive science, they model some psychological theory of a particular aspect of cognition, attempting to account for experimental data. Others aspire to be general computational models of cognition (Sun, 2008). These are usually designed around some unified theory of cognition (Newell, 1990). They include SOAR (Laird, Newell, & Rosenbloom, 1987), ACT-R (Lebiere & Anderson, 1993), CAPS (Just & Carpenter, 1987), Clarion (Sun, 1997), EPAM (Feigenbaum & Simon, 1984), EPIC (Meyer & Kieras, 1997), and LIDA (Franklin & Patterson, 2006; Ramamurthy, Baars, D'Mello, & Franklin, 2006).

These models offer certain tangible benefits to enrich the basic theorize → predict → experiment → theorize cycle described above. In this section we briefly describe some of the advantages of computational modeling along with potential shortcomings. Several of the limitations can be alleviated with a cognitive robotic approach as we will discuss shortly.

1.1. Benefits of computational modeling

Computational models of human cognition that replicate a broad spectrum of experimental data and offer provocative testable hypotheses have two clear benefits. First, *the process of model building is highly conducive to obtaining a deep understanding of the real world phenomenon under consideration*. The model building process involves three phases that compel a researcher to delve deep into the mechanistic core of the phenomenon being modeled (Black, 1962). These include deciding on the functional requirements and overall goal of the model, isolating the various individual components or building blocks of the model, and devising schemes to bring about the interplay between the aforementioned components in order to obtain the desired behavior (e.g. Franklin, 2005; Laird et al., 1987; Lebiere & Anderson, 1993). Each of these phases requires deep thought and bold innovative theories that deepen understanding of the phenomenon being modeled.

The second major benefit of computational modeling lies within the *insights that can be gleaned from basic computational principles that underlie the model*. The underlying assumption is that any design decision made in the process of developing a computational model of some psychological process can be interpreted as a hypothesis that can be tested on human or animal participants. In this fashion, a computational model may indirectly make qualitative predictions that can be experimentally verified.

As an example consider a researcher who decides to endow a cognitive robot with a long term memory (LTM). The first design decision is to select the appropriate computational system that best captures some of the aspects of LTM behavior such as recency effects, rehearsal, interference, etc. One possibility is to design a new computational model for LTM as is done in cognitive architectures such as ACT-R (Lebiere & Anderson, 1993) and SOAR (Laird et al., 1987). As an alternative one may select from one of the several existing associative memory models such as Sparse Distributed Memory (SDM) proposed by Kanerva (1988, 1993), the well known Hopfield model (Hopfield, 1982), or a dynamical systems approach such as the KIII model (Kozma & Freeman, 2001).

The choice of the computational model offers the first testable hypothesis for the psychological model. For example, let us assume that we are testing whether memory retrieval in humans is a convergent or non-convergent process. The SDM model is a non-convergent model while the Hopfield model is a convergent model. If experimental simulations with SDM and the Hopfield models indicate that SDM replicates human data more closely than the Hopfield model, it can be argued that the properties of SDM (non-convergent dynamics) mirror human memory, to greater extent than the Hopfield model (convergent dynamics).

Once a suitable computational model has been selected and implemented, basic experimentation can shed further light into the inner working of the model and the real world phenomenon being modeled. For example, consider the heated debate as to the underlying mechanism behind forgetting in long term memory (LTM). While it is generally accepted that humans forget information, the underlying mechanism (cognitive or neural) that causes the forgetting is debated. Two primary theories and possible mechanisms of forgetting are *decay* (Brown, 1958; Ebbinghaus, 1985/1964; Peterson & Peterson, 1959) and *interference* (Keppel & Underwood, 1962; McGeoch, 1932; Waugh & Norman, 1965). Interference influences forgetting because similar events encoded in a memory system interfere with one another and negatively affect retrieval. Alternately, decay brings about forgetting by causing a loss of memory traces attributed only to time. *Retrieval failures* have also been proposed as the possible basis for forgetting-memories never disappear; they just cannot be retrieved (Tulving, 1968). With a computational model for memory in hand, several experimental simulations can be conducted in order to test each of these positions.

Another powerful advantage of a computational model is that its properties can be established by mathematical analysis, thereby eliminating the need for implementation and experimentation. For example, Keeler (1988) performed a theoretical analysis on the memory capacity of Sparse Distributed Memory and the Hopfield model. On the basis of this analysis a researcher can easily determine whether interference in each model would account for forgetting or whether an explicit decay mechanism would be required.

1.2. Limitations of computational models

The major limitation of computational modeling when attempting to replicate experimental data on humans is that the perception and action capabilities of the various computational models of human cognition are impoverished as these models are not structurally coupled to the real world via sensors and actuators. Simply put, it is difficult to achieve structural validity with computational modeling (Brooks & Tobias, 1996; Frijda, 1967; Zeigler, 1976). In order to achieve structural validity (or strong equivalence; Fodor, 1968), a model must not only produce human-like behavior but also reflect the manner in which humans produce the behavior (Zeigler, 1976). For example, two widely popular models, the Construction-Integration model (Kintsch, 1988) and SOAR (Laird et al., 1987), require propositionalized input (i.e. red(ball), object(chair), etc.) which clearly differs from the manner in which experimental stimuli are presented to human participants in the same experiments.

For example, consider a typical memory experiment where human participants are required to memorize and recall a list of words, or a reading comprehension experiment where an eye tracker is used to monitor eye movements while participants read a passage. If one attempted to replicate such experiments with a computational model, the input and output methods for the model would differ from those employed with the human participants, thereby threatening the validity of the replication. With the memory experiment each word in the list would be inserted into the computational systems' memory. For the reading comprehension task the passage would be fed into the system, either as a whole, or in incremental parts as words and sentences. In contrast, when experimenting with humans, stimuli are presented on a piece of paper or a computer screen, and it is up to the participants to sense, perceive, and represent the information. They could, for example, make a simple reading mistake that the computer could not.

Similar to the problems associated with limited real world sensory processing capabilities, other complications arise in attempts to replicate experiments where the action component is more than simple output to a screen or mouse movements. By eliminating real world action, the scope of human experimental data that is suitable for replication with computational models is limited. For example, experiments involving bodily action such as grasping, navigation, posture, gesture, eye movements, speech modulation, etc., lie beyond the scope of what is possible with classical computational modeling.

Of course, one might argue that tasks involving rich sensory-motor interactions are of less or no importance to the study of human cognition which is still dominated by higher-level processes such as memory, language, comprehension, attention, categorization, problem solving, and reasoning. However, this classical view of human cognition is being challenged on a number of fronts by various situated or embodied theories of cognition (Chiel & Beer, 1997; Clark, 1997; Glenberg, Havas, Becker, & Rinck, 2005; Varela, Thompson, & Rosch, 1991; de Vega, 2002). Theories of embodied cognition postulate that cognitive processes are constrained substantially by the environment and by the coupling of perception and action. Although the validity of embodied theories is currently an issue of serious contention in cognitive science (see de Vega, Glenberg, & Graesser, 2008), the emergence of these theories has further narrowed the bandwidth of possible experimental replications via computational models.

1.3. Cognitive robots can alleviate the limitations of computational modeling

In designing a model for some phenomenon a researcher is free to establish his or her level of abstraction within which to work (Brooks & Tobias, 1996). For example, for the long term memory example discussed above, the various levels of abstractions can be aligned in steps ranging from the model being heavily rooted in neuroscience to a purely functional model. The neural model would attempt to isolate brain regions that participate in the encoding and retrieval of memories, while the functional model would simply describe the behavior of the LTM system under varying conditions. A computational model lies somewhere between these extremes towards a higher level of abstraction. Therefore, though useful in replicating experimental data from humans, it inevitably violates several principles of ecological validity mainly due to the use of artificial environments.

On the other hand, a cognitive robot "lives" in the real world and must adapt in order to survive in complex, dynamic, and sometimes unpredictable environments. Being structurally coupled to the world via its sensors and effectors, a cognitive robot operates towards the low end of the abstraction scale. Therefore, an ecologically valid environment can be obtained with a cognitive robot. Additionally, by participating in an experiment, a cognitive robot can test not only functional predictions, but also the suitability of the underlying mechanisms. For example, a LTM experiment involving a cognitive robot employing sparse distributed memory may provide evidence of the suitability of a distributed memory mechanism that could not be obtained from a computational model that was primarily mathematical in nature. Reaction time experiments with word recognition (e.g. recency effects) assume that the time between the recognition decision and the pressing of the appropriate key is independent of the recency of the word. Such assumptions can be directly tested using a cognitive robot.

2. Psychological modeling to guide the development of a cognitive robot

As designers of intelligent systems, roboticists have always been concerned with what to do and how to do it. In the early days of robotics research emphasis was primarily focused on low-level sensing and control tasks including sensory processing, path planning, and manipulator design and control, that is, with *how* to do it. While the systems developed provide the basis of several fundamental operations of any robot, the intelligence of the robots developed is comparable to that of a standard assembly line robotic system. Today, cognitive robotics is coming online (Dastani et al., 2002; Franklin, 2005). Their emphasis is on cognitive control, that is, with endowing robots with higher-level cognitive functions that enable them to perceive, reason and act in dynamically changing and unpredictable environments. Their primary concern is with choosing *what* to do.

2.1. Design paradigms for a cognitive robot

In this paper we assume that in order to develop a cognitive robot that demonstrates some form of human-like intelligence one should model its control structure (or mind) after humans. However, it is important to note that this is not a necessary assumption. Robots that demonstrate remarkable aspects of intelligent behavior by means of purely computational solutions have been produced with little or no psychological modeling (akin to AI systems that compete with the best chess players in the world). Our position is that adhering to psychological principles based on empirical findings and innovative theories of human cognition in the design of a cognitive robot yields unique benefits over other design paradigms. Before we defend this position, however, it is important to understand what constitutes a cognitive robot.

Although difficult to define, it is generally accepted that a cognitive robot is equipped with a set of higher order cognitive capabilities such as goals, motivations, reasoning, deliberation, etc (Kawamura & Browne, in press). It is by virtue of these facilities that a cognitive robot can be clearly distinguished from traditional robots such as assembly line robots that are rich in sensory-motor interactions within a restrictive environment but fall short of demonstrating higher-level intelligence. Some researchers have adopted a more conservative definition as to what constitutes a cognitive robot. In particular, Clark and Grush (1999) recommend cognitive phenomena to "involve off-line reasoning, vicarious environmental exploration, and the like" (p. 12). Although several other researchers echo a similar view on what constitutes truly cognitive behavior (e.g. Beer, 2003; Harvey, Di Paolo, Quinn, & Wood, 2005; Kelso, 1995; Thelen & Smith, 1994), our view is more in line with embodied theories that state that perception and action are intricately bound to cognition; therefore both bottom up as well as top down processes can be considered to be cognitive (Clark, 1997; Glenberg et al., 2005; de Vega, 2002).

There is also no general consensus or established design paradigm to scaffold the development of a cognitive robot. Simply put, we can define *what* constitutes a cognitive robot but not *how* to go about building one. There are at least two diverse approaches that one can employ towards the development of the cognitive robot: a strict AI approach using a cognitive model based design and the heavily rooted psychological based approach that is advocated here and by some of our colleagues. The AI based design paradigm relies on tried and tested concepts such as knowledge representation, heuristic search, and planning. While this approach can certainly be leveraged to obtain advancements in robotics, its potential as a medium to obtain higher-level intelligence and true understanding is questionable due to its reliance on problematic assumptions such as the symbol systems hypothesis (Brooks, 1986, 1990) rather than on empirical knowledge from cognitive psychology.

An attractive alternative is to base the control structure of a robot on computational models of human cognition. However, additional benefits may be obtained by extending the computational models of cognition paradigm by requiring all elements of the model to be consistent with existing psychological theories of cognition. Examples of such theories include including situated or embodied cognition (Clark, 1997; Glenberg, 1997; Varela et al., 1991), Barsalou's theory of perceptual symbol systems (1999, Harnad, 1990), working memory (Baars & Franklin, 2003; Baddeley, 1992, 2000; Baddeley & Hitch, 1974), Glenberg's theory (1997) of the importance of affordances (Gibson, 1979) to understanding, the longterm working memory of Ericsson and Kinstch (1995); Baars' global workspace theory (1988, 1997).

In our view computational models of human cognition that are consistent with related psychological theories provide the best design paradigm for cognitive robots. This is because such an approach is expected to yield both engineering and scientific gains. We expect engineering improvements because we are basing our computational mechanisms on the best known example of intelligence, i.e. humans. Scientific gains can be achieved by using computer systems to test, and perhaps augment, psychological theories of cognition.

2.2. Computational mechanisms to scaffold the development of a cognitive robot

A cognitive robot embedded within a complex, dynamically changing environment must frequently and cyclically sample (sense) its environment and act on it. Therefore, the concept of a cognitive robot is deeply rooted in a broader class of systems known as autonomous agents. Franklin and Graesser (1997) define an autonomous agent as "a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future".

Following this definition, in order to function in its environment the control structure of a cognitive robot must be sufficiently broad to encompass basic cognitive processes such as perception, episodic memory, selective attention, action selection, and action execution. We consider these processes to be fundamental, each playing a unique but collaborative role in the functioning of a cognitive robot. We briefly describe each of these processes and also list possible computational mechanisms composed of "new AI" techniques that might be useful in modeling each process.

2.2.1. Perception

An integrated perceptual system is essential for any cognitive robot in order for it to recognize, categorize, understand, and integrate information about its world. One possible model of organization would be a semantic net with activation based on Hofstadter and Mitchell's Copycat architecture (1994). Nodes and links of the network constitute the agent's perceptual symbols (Barsalou, 1999) representing individuals, categories, relations, situations, etc (Franklin, 2005).

2.2.2. Episodic memory

Episodic memory is a longterm memory for the *what*, *when*, *and* where of events. Such a memory system would allow a cognitive robot to recall episodes from its past and use this information to influence decision making. A robot might also be endowed with a semantic memory system for long term storage of factual world knowledge.

Sparse Distributed Memory is an attractive computational model for long-term episodic memory because it captures several of the functional characteristics so typical of human episodic memory. These include recency effects, rehearsal, interference, knowing that one knows, tip of the tongue effect, etc (Kanerva, 1988, 1993). Additional benefits are gained from it being content addressable and associative, though other memory systems such as the Hopfield model (Hopfield, 1982) also share these properties.

2.2.3. Working memory

Working memory provides a place and a process for percepts from perceptual memory and local associations from episodic memory to be combined into structures representing the agent's beliefs about its current situation in its world (Baars & Franklin, 2003; Baddeley & Hitch, 1974). Working memory can be implemented in a number of ways as evidenced by the differences in working memory systems in the various cognitive architectures (see Miyake & Shah, 1999).

2.2.4. Selective attention

Due to real world limitations on the amount of knowledge that can be engineered into a cognitive robot, the robot will have to deal with novel or problematic situations that cannot be dealt with efficiently, or at all, by automatized processes. In these situations, selective attention provides access to appropriately useful internal resources, thereby solving the *relevance* problem, that is, the problem of identifying those internal resources that are relevant to the current situation. Global Workspace Theory (Baars, 1988, 1997) and its fleshing out in the LIDA model (Ramamurthy et al., 2006) provide unique insights into how a selective attentional mechanism may be incorporated into computational systems.

2.2.5. Procedural memory

The schema mechanism developed by Drescher (1991) can be used to computationally model procedural memory. Built-in primitive (empty) schemas directly controlling effectors are analogous to motor cell assemblies controlling muscle groups in humans. A schema consists of an action, together with its context and its result. The memory system organizes schemas into parallel groups for simultaneous action and ordered sequences for sequential processing.

2.2.6. Action selection

Deciding "what to do next" is essential for any agent be it human or a cognitive robot (Franklin, 1995). Simply put, a cognitive robot must utilize the information it perceives in order to select an action in service of its goals. Maes' (1989) behavior net provides for high-level action selection that effectively balances being goal oriented, opportunistic, and thoughtful.

2.2.7. Putting it all together

A model based on several specialized mechanisms, each implementing various facets of human cognition, requires an iterative process to bring about the functional interaction among the various components. There are several ways in which this can be accomplished. One might adhere to a SOAR based design paradigm where items perceived are utilized to recall associations and recruit resources from procedural memory in several knowledge elaboration cycles (Laird et al., 1987). If conflicts arise, a decision cycle is initiated, and the chosen action is executed. Alternatively, one might follow the ACT-R model where each component interacts with the others through limited capacity buffers (Lebiere & Anderson, 1993). The LIDA model may be chosen where selective attention plays a critical role in the recruitment of resources and in various types of learning (Baars & Franklin, 2003; Franklin, Baars, Ramamurthy, & Ventura, 2005; Ramamurthy et al., 2006). With the basic components in place, a cognitive robot adhering to each of these design principles can be programmed and deployed. With each type of model integration method offering a unique hypothesis for similar functionality in humans, the one that accounts for the most experimental data might be preserved.

3. Cognitive developmental robots

For cognitive robots immersed in simple, mostly static domains, it is quite plausible for the system designer to embed the required knowledge so that the agent can effectively pursue its agenda. This approach, though tedious in it's undertaking, has worked quite successfully for knowledge based systems such as expert systems. However, as the complexity of the robot's world increases, the requisite knowledge engineering proves to be an extremely daunting task. In an attempt to circumvent these knowledge engineering problems, we argue for mechanisms that support a developmental period, one of rapid learning in the "lives" of the cognitive robots. Such a developmental period would circumvent the necessity of designing and implementing a complex ontology, a clear pragmatic advantage. In complex, dynamic environments, the learned ontology can be expected to out perform one designed and built-in with much less human effort.

A cognitive developmental robot (Asada, MacDorman, Ishiguro, & Kuniyoshi, 2001; Lungarella, Metta, Pfeifer, & Sandini, 2003; Zlatev & Balkenius, 2001) needs robust, flexible learning mechanisms to acquire relevant world knowledge and to organize this knowledge to facilitate interaction with its world via perception and action. We can identify four fundamental types of learning that would be essential for a cognitive developmental robot. These include: 1) perceptual learning, the learning of new objects, categories, relations, etc.; 2) episodic learning of events, the what, where, and when; 3) procedural learning, the learning of new actions and action sequences with which to accomplish new tasks; and 4) attentional learning, learning what to attend to, or "learning to attend." (Gelman, 1969).

Similar to our arguments for basing the design principles for a cognitive robot on psychological theories of human cognition, there are unique advantages to basing the learning mechanism of a developmental robot from principles of human-like learning. But first let us examine some of the limitations of AI based machine learning techniques in order to explore avenues where psychological principles may guide the development of a cognitive developmental robot.

Dating back to Samuel's (1959) checkers player, machine learning is among the oldest of the sub-branches of AI with many practitioners and many successes to its credit. Still, after fifty years of effort there are remaining difficulties. Machine learning often requires large, accurate training sets, shows little awareness of what's known or not known, integrates new knowledge poorly into old, learns only one task at a time, allows little transfer of learned knowledge to new tasks, and is poor at learning from human teachers.

In contrast, human learning has solved many of these problems and is typically continual, quick, efficient, accurate, robust, flexible, and effortless. For example, consider perceptual learning, the learning of new objects, categories, relations, etc. Traditional machine learning approaches such as object detection, classification, clustering, etc., are highly susceptible to the problems raised above. However, perceptual learning in humans and animals seems to have no such restrictions. Perceptual learning in humans occurs incrementally so there is no need for a large training set. Learning and knowledge extraction are achieved simultaneously through a dynamical system that can adapt to changes in the nature of the stimuli perceived in the environment. Additionally, human-like learning is based on reinforcement rather than fitting to a data set or model. In addition to learning, humans can also forget. Initially, many associations are made between entities. The ones that are sufficiently reinforced persist, while the ones that are not decay away.

There are several benefits to basing the learning mechanisms of a cognitive developmental robot on principles of human-like learning. No large training sets would be required. New knowledge would be easily integrated into old. Several tasks could be learned concurrently with transfer of knowledge to new tasks. With a cognitive developmental robot we can begin to replicate experimental data associated with learning in humans, thereby contributing to one of the main goals of cognitive science.

4. Feasibility of cognitive robotics

The most obvious disadvantage of cognitive robotics is the significant research investment and substantial engineering costs associated with producing these robots. However, as the technology matures, prices continue to drop and an increasing variety of robotic systems are now available for purchase at moderate to steep prices. Additionally, robotic simulation environments offer the best alternative for researchers in search of a cost-efficient environment for rapid development and experimentation.

Simulation environments provide a researcher with a sensory-rich world with a high degree of user control in order to test ideas in a relatively inexpensive manner. To use a simulator, a researcher embeds a control structure that is consistent with the senses and effectors of the robot being tested. In this fashion, experiments that include several configurations of control algorithms for different robots can be easily conducted.

Among the more popular robotic simulators are Webots, Usarsim, Mobilesim, and Gazebo. These simulators share several common features. First, each provides a detailed model of a variety of real robotic systems. Specifically, the physical characteristics of a robot such as weight, motor speeds, and sensory inputs are modeled at a sufficient level of detail. For example, consider a robot with an infrared sensor. The simulator would provide a range of readings based on the distance between the simulated robot and an obstacle that have been appropriately calibrated for the real robot.

Each simulator also provides several default "worlds" or "environments" for a robot to "live" in. These are usually customizable and new environments can be easily created using simple markup languages. The robot being simulated now "lives" in this world, can navigate through the world, senses objects in the world, and can sometimes interact with these objects, thereby demonstrating the essential qualities for an autonomous robotic agent.

As a concrete example, consider Webots, a robotic simulator package for several varieties of wheeled, legged and flying robots. This software package supports several popular commercial robots such as E-Puck, Khepera, IPR, Aibo, and Lego. Additionally, these robots are simulated with excruciating detail. At a minimal level of abstraction eight Infra Red sensors and two servo motors can be simulated for the Khepera robot. Optional features such as a gripper and a camera can also be easily simulated. Please see Dozier (2001), Ivica and Kerstin (2003), and Kerstin and Coles (2001) for some innovative examples that illustrate the use of an external control structure to control a Khepera robot within the Webots simulation environment.

5. Discussion

We conclude by highlighting the usefulness of cognitive robotics towards obtaining a better understanding of human cognition. We believe that large scale working models of cognition that are mechanistic embodiments of functional psychological theories can be useful tools to guide the research of psychologists and cognitive scientists by generating testable hypotheses about human cognition and by providing the means of testing such hypotheses empirically. Such a cognitive robot would generate hypotheses about human cognition by way of its design, the mechanisms of its modules, their interaction, and its performance in either real or simulated environments. In principle, all of these hypotheses are testable; however, due to the relatively fine-grained level of analyses required to either confirm or reject them, more sophisticated brain and behavioral assessment technologies may be in order.

Designing cognitive robots to perform realworld tasks will produce working models of cognition. Though experiments provide the gold standard for scientific evidence, it is not possible to test all parameters of actual working models of cognition. Experiment-based models typically have too few variables to accomplish realworld perception or control of action. Simulations based only on experimental evidence would simply fail in the real world. Hence, workability should be combined with experimental evidence as desirable features of cognitive models. The reciprocal interactions we have described between computational modeling using cognitive robots and functional psychological experimentation is a methodological change that should make a significant impact in theoretical and experimental psychology. In particular, it should allow experimenters to deal with concepts at lower levels of abstraction, to factor in the effects of perception and action, and to make distinctions at minute time scales. In addition, computational design decisions will yield testable hypotheses, while computational principles and mechanisms can guide both psychologists and neuroscientists in looking for psychological and neural mechanisms that underlie cognitive processes. Similar to the use of robots as models of biological behavior of a variety of organisms such as lobsters (Grasso, Consi, Mountain, & Atema, 1996, 2000), paper wasps (Honma, 1996), cockroach kinematics (Quinn

& Ritzmann, 1998), rat hippocampus (Burgess, Donnett, & O'Keefe, 1998), and others (see Beer, Chiel, Quinn, & Ritzmann, 1998; Webb, 2001), when used in conjunction with psychological experimentation, cognitive robotics should prove to be especially beneficial to cognitive science.

References

Angell, J. (1907). The Province of functional psychology. *Psychological Review, 14*, 61–91.

Asada, M., MacDorman, K. F., Ishiguro, H., & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems, 37*, 185–193.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. J. (1997). *In the theater of consciousness*. Oxford: Oxford University Press.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Science, 7*, 166–172.

Baddeley, A. D. (1992). Consciousness and working memory. *Consciousness and Cognition, 1*, 3–6.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science, 4*, 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Barto, A. G. (1991). Learning and incremental dynamic programming. *Behavioral and Brain Sciences, 14*, 94–95.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior, 11*(4), 209–243.

Beer, R. D., Chiel, H. J., Quinn, R. D., & Ritzmann, R. E. (1998). Biorobotic approaches to the study of motor systems. *Current Opinion in Neurobiology, 8*, 777–782.

Beveridge, W. I. B. (1957). *The art of scientific investigation*. Vintage/Alfred A. Knopf.

Block, N. (1980)*Readings in the philosophy of psychology, Vols. 1 and 2*. Cambridge, MA: Harvard University Press.

Bower, J. M. (1992). Modeling the nervous system. *Trends in Neurosciences, 15*, 411–412.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, RA-2*(1), 14–23.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems, 6*, 3–15.

Brooks, R. J., & Tobias, A. M. (1996). Choosing the best model: level of detail, complexity and model performance. *Mathematical Computer Modelling, 24*, 1–14.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12–21.

Burgess, N., Donnett, J. G., & O'Keefe, J. (1998). Using a mobile robot to test a model of the rat hippocampus. *Connection Science, 10*, 291–300.

Chiel, H., & Beer, R. (1997). The brain has a body: adaptive behaviour emerges from interactions of nervous system, body and environment. *Trends in Neurosciences, 20*, 553–557.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Clark, A., & Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior, 7*(1), 5–16.

Dastani, M., de Boer, F., Dignum, F., van der Hoek, W., Kroese, M., & Meyer, J. J. (2002). Programming the deliberation cycle of cognitive robots. In Proceedings of the third international cognitive robotics workshop. Edmonton, Canada.

Dozier, D. (2001). Evolving robot behavior via interactive evolutionary computation: from real world to simulation, Symposium on applied computing. Proceedings of the 2001 ACM symposium on applied computing.

Drescher, G. (1991). *Made up minds: A constructivist approach to artificial intelligence*. Cambridge, MA: MIT Press.

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. New York: Dover.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305–336.

Fodor, J. A. (1968). *Psychological explanation*. Random House.

Franklin, S. (1995). *Artificial minds*. Cambridge MA: MIT Press.

Franklin, S. (2005). Cognitive robots: perceptual associative memory and learning. In Proceedings of the 14th Annual International Workshop on Robot and Human Interactive Communication (RO-MAN 2005), pp. 427–433.

Franklin, S., Baars, B. J., Ramamurthy, U., & Ventura, M. (2005). The role of consciousness in memory. *Brains, Minds and Media, 1*, 1–38.

Franklin, S., & Graesser, A. C. (1997). Is it an agent, or just a program?: a taxonomy for autonomous agents. Proceedings of the Third international workshop on agent theories, architectures, and languages, published as intelligent agents III (pp. 21–35), Springer-Verlag.

Franklin, S., & Patterson, F. G. (2006). *The Lida architecture: Adding new modes of learning to an intelligent, autonomous, software agent*. Integrated Design and Process Technology, IDPT-2006. San Diego, CA: Society for Design and Process Science.

Frijda, N. J. (1967). Problems of computer simulation. *Behavioural Science, 12*, 59–67.

Gelman, R. S. (1969). Conservation acquisition: a problem of learning to attend to relevant attributes. *Journal of Experimental Child Psychology, 7*, 167–187.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1–19.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states: the case for emotion. In R. Zwaan, & D. Pecher (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.

Grasso, F., Consi, T., Mountain, D., & Atema, J. (1996). Locating odor sources in turbulence with a lobster inspired robot. In P. Maes, M. J. Mataric, J. A. Meyer, J. Pollack, & S. W. Wilson (Eds.), *Sixth international conference on simulation of adaptive behaviour: From animals to animats 4*. Cambridge, MA: MIT Press.

Grasso, F., Consi, T., Mountain, D., & Atema, J. (2000). Biomimetic robot lobster performs chemo-orientation in turbulence using a pair of spatially separated sensors: progress and challenges. *Robotics and Autonomous Systems, 30*, 115–131.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Harvey, I., Di Paolo, E. A., Quinn, M., & Wood, R. (2005). Evolutionary robotics: a new scientific tool for studying cognition. *Artificial Life, 11*(1–2), 79–98.

Hofstadter, D. R., & Mitchell, M. (1994). The copycat project: a model of mental fluidity and analogy-making. In K. J. Holyoak, & J. A. Barnden. (Eds.), *Logical connections. Advances in connectionist and neural computation theory, Vol. 2*. Norwood, NJ: Ablex.

Honma, A. (1996). Construction robot for three-dimensional shapes based on the nesting behavior of paper wasps. *Seimitsu Kogaku Kaishi/Journal of the Japan Society for Precision Engineering, 62*, 805–809.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences of the USA, 79*, 2554–2588.

Ivica, M., & Kerstin, D. (2003). Social attitudes: investigations with agent simulations using webots. *Journal of Artificial Societies and Social Simulation, 6*(4).

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.

Kanerva, P. (1988). *Sparse distributed memory*. Cambridge, MA: The MIT Press.

Kanerva, P. (1993). Sparse distributed memory and related models. In M. H. Hassoun (Ed.), *Associative neural memories: Theory and implementation*. New York: Oxford University Press.

Kawamura, K., & Browne, W. N. Cognitive robotics. In Robert A. Meyers (Ed.), Springer encyclopedia of complexity and systems science. Heidelberg, Germany: Springer Science, (in press).

Keeler, J. D. (1988). Comparison between Kanerva's SDM and Hopfield-type neural networks. *Cognitive Science, 12*(3), 299–329.

Kelso, J. (1995). *Dynamic patterns*. Cambridge, MA: MIT Press.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153–161.

Kerstin, D., & Coles, S. J. (2001). Narrative intelligence from the bottom up: a computational framework for the study of story-telling in autonomous agents. *Journal of Artificial Societies and Social Simulation, 4*(1).

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*(2), 163–182.

Kozma, R., & Freeman, W. J. (2001). Chaotic resonance – methods and applications for robust classification of noisy and variable patterns. *International Journal of Bifurcation & Chaos, 11*, 1607–1629.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artificial Intelligence, 33*, 1–64.

Lebiere, C., & Anderson, J. R. (1993). A connectionist implementation of the ACT-R production system. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society. Hillsdalle, NJ: Erlbaum.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109*, 376–400.

Losee, J. (1980). *A historical introduction to the philosophy of science* (2nd ed.). Oxford: Oxford University Press.

Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: a survey. *Connection Science, 15*, 151–190.

Maes, P. (1989). How to do the right thing. *Connection Science, 1*, 291–323.

Maynard Smith, J. (1974). *Models in ecology*. Cambridge: Cambridge University Press.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*, 352–370.

Meissner, W. W. (1966). The implications of experience for psychological theory. *Philosophy and Phenomenological Research, 26*, 503–528.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance, Part 1: basic mechanisms. *Psychological Review, 104*, 3–65.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.

Quinn, R. D., & Ritzmann, R. E. (1998). Construction of a hexapod robot with cockroach kinematics benefits both robotics and biology. *Connection Science, 10*, 239–254.

Ramamurthy, U., Baars, B. J., D'Mello, S. K., & Franklin, S. (2006). LIDA: a working model of cognition. In D. Fum, F. Del Missier, & A. Stocco (Eds.), Proceedings of the 7th international conference on cognitive modeling (244–249). Trieste: Edizioni Goliardiche.

Salmon, W. C. (1990). *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*, 210–229.

Sun, R. (1997). An agent architecture for on-line learning of procedural and declarative knowledge. In Proceedings of the International Conference on Neural Information Processing (ICONIP'97): Progress in Connectionist-Based Information Systems. Singapore: Springer Verlag.

Sun, R. (Ed.). (2008). *The Cambridge handbook of computational psychology*. New York: Cambridge University Press.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.

Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon, & D. L. Horto (Eds.), *Verbal behaviour and general behaviour theory*. Englewood Cliffs: Prentice Hall.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

de Vega, M. (2002). Del significado simbólico al significado corpóreo. *Estudios de Psicología, 23*, 153–174, [From symbolic meaning to embodied meaning].

de Vega, M., Glenberg, A. M., & Graesser, A. C. (Eds.). (2008). *Symbols, embodiment, and meaning*. Oxford: Oxford University Press.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72*(2), 89–104.

Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioural and Brain Sciences, 24*(6), 1033–1050.

Zeigler, B. P. (1976). *Theory of modelling and simulation*. New York: John Wiley.

Zlatev, J., & Balkenius, C. (2001). Introduction: why Epigenetic Robotics? In Balkenius, C., Zlatev, J., Kozima, H., Dautenhahn, K., & Breazeal, C. (Eds.), Proceedings of the first international workshop on epigenetic robotics: Modeling cognitive development in robotic systems Vol. 85 (pp. 1–4). Lund University Cognitive Studies.