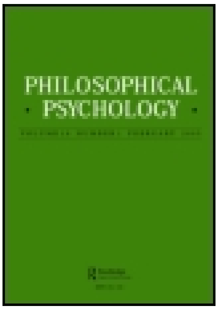

Philosophical Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cphp20

Autonomous generation of symbolic representations through subsymbolic activities

Ron Sun

Version of record first published: 04 Sep 2012.

Philosophical Psychology
2012, 1–25, iFirst

Autonomous generation of symbolic representations through subsymbolic activities

Ron Sun

This paper explores an approach for autonomous generation of symbolic representations from an agent's subsymbolic activities within the agent-environment interaction. The paper describes a psychologically plausible general framework and its various methods for autonomously creating symbolic representations. The symbol generation is accomplished within, and is intrinsic to, a generic and comprehensive cognitive architecture for capturing a wide variety of psychological processes (namely, CLARION). This work points to ways of obtaining more psychologically/cognitively realistic symbolic and subsymbolic representations within the framework of a cognitive architecture, and accentuates the relevance of such an approach to cognitive science and psychology.

Keywords: Bottom-Up Learning; Cognitive Architecture; Psychology; Rule Extraction; Symbol Grounding

1. Introduction

Some of the fundamental questions concerning human cognition/psychology, as addressed by cognitive science, include: how are concepts and their corresponding symbolic representations generated and maintained? How are such items (i.e., concepts; in particular, symbolic representations of concepts) organized in memory? How does one comprehend and produce symbolic structures? How does one construct or re-construct the meaning of such structures? How are one's symbolic representations related to the world?

The central matter underlying these questions is the generation of symbolic representations and how they are linked to other mental content and to the world

Ron Sun is Professor of Cognitive Sciences at Rensselaer Polytechnic Institute.
Correspondence to: Ron Sun, Cognitive Sciences Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA. Email: rsun@rpi.edu

ISSN 0951-5089 (print)/ISSN 1465-394X (online)/12/000001-25 © 2012 Taylor & Francis
http://dx.doi.org/10.1080/09515089.2012.711035

(Sun, 2000); for these questions necessarily involve the generation of symbolic representations and their interpretations (among other aspects).

The focus of the present paper will specifically be on autonomous but interactive generation of symbolic representations within the context of everyday agent-world interaction. It is autonomous in the sense that a cognitive agent generates symbolic structures without being given (literally) such structures from the external sources (Weng et al., 2001). It is interactive in the sense that the generated symbolic structures are the results of embodied/situated interaction between the agent and the world, and their ontogenetic, sociocultural, and even phylogenetic and evolutionary history (Varela, Thompson, & Rosch, 1991). Through the use of a computational cognitive architecture, this process is made computationally (i.e., in a process-based, mechanistic sense) possible.

In the remainder of this paper, first, some background issues concerning situated interactions and symbolic processing are discussed (in section 2). Then, specific ways of situated interactions within the CLARION cognitive architecture are sketched (in section 3). In section 4, details of how symbols are generated within CLARION are presented. Some generic examples follow in section 5. Section 6 provides further discussions that draw out the implications of the approach discussed. Section 7 concludes the paper.

2. Background and Issues

In order to address these issues above, we need to look into the human cognitive architecture; that is, the essential structures and processes of the human mind (to the extent that we can understand or hypothesize about them).

2.1. Some Desiderata

In exploring the human cognitive architecture, a number of essential desiderata may need to be taken into consideration (among others; as argued at length in Sun, 2002):

> **Reactivity.** In human everyday activities, behaviors are mostly generated without involving elaborate computation. Behaviors are often direct and immediate; they are often ''non-representational,'' that is, without involving complex symbolic representations and manipulations (Dewey, 1958; Dreyfus, 1992).

> **Routineness.** Human everyday activities are very much routinized and thus largely made of habitual sequences of behaviors (Sun, 2002). We may view different routines as different sets of skills for coping with the everyday world. Gradual adaptation (or learning) of these routines or skills is important. Overall, human activities may be viewed as composed of forming, changing, and following such routines (Dreyfus, 1992; Sun, 2002).

> **Trial-and-Error adaptation.** The learning of routines is mostly, and essentially, a trial-and-error adaptation process. Various manifestations of such adaptation have been studied in psychology under the rubric of law of effect, classical and instrumental conditioning, and probability learning. This type of learning appears the most essential to the human mind (see, e.g., Reber, 1989).

Second, to explore the inner working of the human mind, some essential cognitive characteristics may be identified (which have been extensively argued for and justified in Sun, 2002; Sun, Merrill, & Peterson, 2001; Sun, Slusarz, & Terry, 2005):

Dichotomy of implicit and explicit processes. Implicit processes are consciously inaccessible (at least not directly), "holistic," and imprecise, but deal better with more complex situations, while explicit processes are consciously accessible and more precise but require more attentional resources (Dreyfus & Dreyfus, 1987; Reber, 1989). Implicit processes are mostly responsible for reactivity, routines, and trial-and-error adaptation. The implicit-explicit dichotomy is closely related to some other dichotomies: for example, the dichotomy of symbolic versus subsymbolic processing, whereby implicit cognition involves mostly subsymbolic processing and explicit cognition involves more symbolic processing (Sun, 2002). It can be justified psychologically by the voluminous empirical studies of implicit and explicit learning, implicit and explicit memory, implicit and explicit perception, and so on (Sun, 2002). These empirical studies lead to justifications for the general distinction between implicit and explicit cognition (roughly, the unconscious and the conscious). The combination of the two types of processes has the potential advantages of synergy (as demonstrated in Sun et al., 2005).

Bottom-Up and top-down learning. The interaction between the two sides of the dichotomy in relation to learning includes top-down (explicit learning first and implicit later), bottom-up (implicit learning first and explicit later), and parallel learning (simultaneous implicit and explicit learning). However, bottom-up learning may be most essential to everyday reactive routine activities (Sun et al., 2001). There are various indications of bottom-up learning, including (1) philosophical arguments, such as Heidegger (1927/1962) and Dewey (1958), in which the primacy of direct interaction with the world (in a mostly implicit way) is emphasized, and (2) psychological evidence of the acquisition and the delayed explication of implicit knowledge (e.g., Karmiloff-Smith, 1986; Sun et al., 2001).

Role of motivation. The essential motives/needs of an agent arise prior to cognition. Such motivations are the foundation of an agent's action, routines, and learning (Sun, 2009; Toates, 1986). In a way, cognition (including learning) has evolved to serve the essential motives/needs of an agent. Cognition, in helping to satisfy needs, has to take into account the regularities and structures of the world. Thus, cognition bridges the needs/motives of an agent and its world (Sun, 2009).

2.2. Situated Cognition

What underlies these desiderata is a particularly acute theoretical concern with the everyday existential context in relation to human cognition. Such a concern can be readily traced back to the phenomenological philosophers such as Heidergger and Merleau-Ponty. Among the diverse notions employed by these philosophers, what is particularly relevant is the notion of "being-in-the-world" (Heidergger, 1927/1962): one's existence in the world, or being-in-the-world, is fundamental to being what one is. Being-in-the-World also entails that individuals constantly interact with the world in an immediate and non-reflective way, as in most of their everyday activities.

This non-reflective way of directly dealing with the world should be emphasized in exploring human psychology. Heidegger (1927/1962) proposed that there is a primordial kind of *comportment* that directly involves an agent with the world, without the mediation of explicit, symbolic representation/computation. Comportment is direct and unmediated; in other words, comportment does not necessarily involve, or presuppose, explicit symbolic representations. To the contrary, such representations presuppose the process of comportment. Direct unmediated comportment is in fact the condition of possibility of all representations. Comportment, according to Heidegger, "makes possible every intentional relation to beings" (1978/1984, p. 135) and "precedes every possible mode of activity in general" (1978/1984, p. 183), prior to explicit knowledge and explicit thinking. That is, comportment is primary. In the everyday world, individuals attune themselves to the world in a primarily implicit, "primordial" way.

It is also believed that such "mindless" everyday interaction and activities (or "coping with the world"; Dreyfus, 1992), on top of biological pre-endowment, is the basis of high-level conceptual (symbolic) thinking and intentionality; that is, it constitutes the content and the meaning of symbolic representations (Sun, 2000). Reflective, deliberative, explicit thinking (involving symbolic representation) occurs only in more unusual circumstances. According to Heidegger, deliberative thinking (when it occurs) presuppose a background of common, everyday practices (Dreyfus, 1992). The background of everyday practices is not represented, in an explicit and elaborate (symbolic) fashion (such as in a rule base of an expert system, which spells out every detail and every minute variation of every possible situation), but is assumed in comportment toward the world. In other words, the most fundamental part of the mind is embodied, not explicitly represented, and thus it is not (directly) accessible (Sun, 2000, 2002).

An important implication of the foregoing discussion is that we need to deal with (what has been commonly referred as) "situated/enactive cognition": that is, cognition that is based on directly interacting with the world (physical or social; e.g., Suchman, 1987; Varela et al., 1991). Regardless of whether one believes that situated/enactive cognition is the most important kind of cognition or not, or even whether it is the only kind worth considering or not (as some would suggest), it is evident that any serious student of cognition must consider this kind of cognition, as well as its role in the overall functioning of agents.

2.3. Symbolic Representation within Situated Cognition

On the other hand, there is another, complementary aspect of cognition. Although some advocates of situated/enactive cognition downplay or even deny the role of symbolic representations in (situated/enactive) cognition, there are sufficient reasons to believe that symbolic representations, in their various forms, are important to agents, even with regard to situated/enactive cognition. This is because (1) psychologically they are demonstrably important, and (2) computationally they are important as well.

In psychology, deductive reasoning, categorization, concept learning, and the like have been shown to result, at least in part, from symbolic representations and symbol manipulation (see, e.g., Bruner, Goodnow, & Austin, 1956; Johnson-Laird & Yang, 2008). Although evidently not all psychological processes involve symbols, symbolic representations and symbol manipulation are nevertheless an indispensable part of cognition, involved (at least in part) in many types of cognitive processes (Sun, 1994). When we throw out the bathwater of the "Physical Symbol Hypothesis," we do not need to throw out the baby (the cognitive roles played by symbols) at the same time.

Symbolic representations are also important in accounting for the theoretical distinction between implicit and explicit psychological processes (roughly, unconscious and conscious processes; Reber, 1989), and, in particular, explicit knowledge with regard to that distinction. The distinction between implicit and explicit processes has been amply demonstrated (albeit not uncontroversially). Given the distinction, there is the need to explain the explicit knowledge that humans do possess as empirically demonstrated (e.g., what they express when they verbalize). Symbolic representations provide the best account by providing the representational substrate for explicit knowledge (e.g., as shown by Sun, 2002). As has been demonstrated before, the representational difference between symbolic and subsymbolic representations may account for the phenomenological and psychological differences between the two types (see, e.g., Helie & Sun, 2010; Sun et al., 2005). Thus, both symbolic and subsymbolic representations are needed in order to explain the full range of cognition.

An important notion here is "dual-representation." In Sun (1994), the following hypothesis was put forth:

> It is assumed in this work that cognitive processes are carried out in two distinct "levels" with qualitatively different mechanisms. Each level encodes a fairly complete set of knowledge for its processing, and the coverage of the two sets of knowledge encoded by the two levels overlaps substantially. (1994, p. 44*)*

The two levels often encode similar, comparable content. But they encode their content in different ways (e.g., symbolic versus subsybmolic). Therefore, they could conceivably utilize different mechanisms. They can thus have qualitatively different "flavors." One reason for having the two levels is that these different levels can potentially work together synergistically, complementing each other (as demonstrated in Helie & Sun, 2010; Sun, 1994; Sun et al., 2005).

Symbolic representations are also important computationally. Artificial intelligence has traditionally based their methodologies on symbolic representation and symbol manipulation. The Physical Symbol Hypothesis pushed this approach to its logical extreme. Even though this approach has been disputed by advocates of situated/enactive cognition as well as by connectionists (e.g., Sun, 2002; Varela et al., 1991), a weaker version focusing instead on the relevance and involvement of symbol manipulation in cognition can still be valid. For instance, symbolic AI has shown, computationally, how much cognition, even though maybe not all or most, can be attributed to such symbol manipulation.

Now the question is: how exactly does an agent (in situated/enactive interaction with the world) generate symbolic representations? Some rough ideas in this regard have been sketched out by a number of researchers in the past. Lakoff argues that "meaningful conceptual structures arise from two sources: (1) from the structured nature of bodily and social experience and (2) from our innate capacity to imaginatively project from certain well-structured aspects of bodily and interactional experience to abstract conceptual structures" (1988, p. 121). Dreyfus has further argued that the problem is how to incorporate an account of how we "directly pick up significance and improve our sensitivity to relevance" (2007, p. 265), and that this ability depends on our responding to what is significant to us, given the context. Johnson suggests "the idea that understanding is an event in which one has a world, or more properly, a series of ongoing related meaning events in which one's world stands forth" (1987, p. 175). The upshot is that symbolic representations (and their meanings) stand out from background activities during and through the interaction between the agent and the world. It is not just about receiving information from the world but also projection onto the world. The key issue here is the roles of embodiment, interaction, context, function, need, and goal.

In this regard, it should be emphasized that the world is often seen (constructed) through the existential context of an agent. Perception, action, and sense-making constitute a process laden with the necessary existential concerns, due to the existence of a distinct self, which is in a precarious position and therefore needs to establish valuation in terms of whether situations, events, or actions are conducive or detrimental to its survival (Di Paolo, 2005). An agent treats the situations that it encounters often (if not always) in relation to the viability and functioning of the self. Therefore, these situations may obtain meanings relative to the needs, goals, and actions of the agent (Masmoudi, Dai, & Naceur, 2012; Sun, 2009).

This is the basis of sense-making. The significance that is brought forth by the activity and experience of the agent is what makes the world the way it is (that is, the way it appears to the agent). Meaning is not to be found in the world alone or in the internal dynamics of the agent alone, but in their interaction (Merleau-Ponty, 1963). However, this does not entail that meaning can be reduced to relational phenomena: there is an asymmetry in the relationship between the two, because the relationship is enacted by the activity of the agent (and thus resulting from the goals and needs of the agent).

It is important to emphasize once again the notion of "bottom-up learning," as alluded to earlier (Sun et al., 2001). Bottom-Up learning, not just the psychological notion but also its associated mechanistic (computational) details (Sun, 2002, 2003), provides a concrete way for the sense-making in the afore-discussed manner. As sketched earlier, "senses" (in the form of symbolic, conceptual representations) "pop" out of the background of on-going activities (which, in contrast, are mostly in the form of subsymbolic activities). That is, in our terminology, implicit processes give rise to explicit processes, in the midst of the on-going interaction between the agent and its world in relation to its goals and needs. This is the process of bottom-up learning as proposed in Sun et al. (2001).

3. Symbolic and Subsymbolic Representations in CLARION

To substantiate the ideas discussed above and to fit various pieces together to form a coherent (theoretical and computational) whole, let us look into a psychologically plausible computational cognitive architecture—CLARION (Sun 2002, 2003).

3.1. Overall Approach of CLARION

Let me sketch briefly a general picture of the theoretical approach behind the CLARION cognitive architecture, and what is essential for an integrative cognitive architecture (e.g., situated versus reflective, subsymbolic versus symbolic, and other components).

This approach starts small: there will be only minimum initial structures. Some of these initial structures have to do with behavioral predispositions (e.g., evolutionarily pre-wired reflexes, or predispositions for developing such reflexes); that is, genetic and biological pre-endowment. Some others have to do with learning capabilities: most of the structures and contents of cognition will have to be "constructed" (learned) in an incremental fashion during the course of individual ontogenesis. The development of structures and contents is through interacting with the world (i.e., through "being-in-the-world"), which includes both the physical world and the sociocultural world. The interaction leads to the formation of various behavioral routines, which in turn lead to the emergence of higher-level, symbolic, conceptual representations.

The generation of higher-level mental structures and contents (i.e., higher-level conceptual representations with symbols), on the other hand, is (to a significantly extent) determined by lower-level (subsymbolic) structures and contents (i.e., the idea of bottom-up learning, as sketched before). There is also another source: sociocultural influence, especially through symbols existing and employed in a given culture. Culture also has the role of structuring (constraining) the interaction of an agent with the world through mediating tools, signs, and other cultural artifacts, and thus culture affects lower-level behaviors too (although to a lesser extent maybe).

On this view, higher-level symbolic conceptual representation is rooted in lower-level processes, from which it obtains its meaning and for which it provides focus and clarity. The rootedness is guaranteed by the way higher-level symbolic representations are produced—they are (in the main) extracted out of the lower-level behavioral structures and contents (through bottom-up learning). Even culturally transmitted symbols have to be linked up, within the mind of an agent, with lower-level processes in order to be effective.

This approach has been embodied in the psychologically realistic computational cognitive architecture CLARION, which will be described below (for detailed psychological evidence in support of CLARION, see, e.g., Sun et al., 2001, 2005).

3.2. Structure of CLARION

CLARION consists of a number of distinct subsystems, with a dual representational structure in each subsystem. Its subsystems include the action-centered subsystem (the ACS), the non-action-centered subsystem (the NACS), the motivational subsystem (the MS), and the meta-cognitive subsystem (the MCS). The role of the action-centered subsystem is to control actions (i.e., to maintain and apply procedural knowledge, as is known in psychology), regardless of whether the actions are for external physical movements or for internal mental operations. The role of the non-action-centered subsystem is to maintain general knowledge (i.e., declarative knowledge, as is known in psychology), likely in the service of action decision making by the ACS. The role of the motivational subsystem is to provide underlying motivations for perception, action, and cognition, in terms of providing impetus and feedback. The role of the meta-cognitive subsystem is to monitor, direct, and modify the operations of the other subsystems for better performance (Sun, 2009).

Each of these interacting subsystems consists of two levels of representations (i.e., a dual representational structure): generally, in each subsystem, the top level encodes explicit knowledge with symbolic representations, and the bottom level encodes implicit knowledge with subsymbolic representations. (The distinction of implicit and explicit knowledge has been well argued for; for its psychological origin, see, e.g., Reber, 1989; for further theoretical elucidation, see Sun, 2002; 2012.) The two levels interact, for example, by cooperating in actions, as well as by cooperating in learning (through a bottom-up and a top-down process, to be detailed below). CLARION is thus a dual-representational, dual-process theory of mind. Furthermore, it is based on a four-way partitioning of procedural versus declarative and implicit versus explicit knowledge/processes (see Sun, 2012 for theoretical and empirical arguments for the four-way partitioning). In CLARION, these different types of knowledge arise from different but interacting learning mechanisms (Sun, 2002, 2003). See Figure 1 for a sketch.

3.3. Some Details of CLARION

3.3.1. The action-centered subsystem

First, in the action-centered subsystem of CLARION, the process for action decision making is essentially the following:

> Observing the current state of the world (the observable input state), the two levels of processes within the ACS (implicit or explicit) make their separate decisions in accordance with their respective knowledge, and their outcomes are integrated. Thus, a final selection of an action is made, and the action is then performed. The action changes the world. Comparing the changed input state with the previous input state, the agent learns (e.g., in accordance with Q-learning; more later). The cycle then repeats itself.

In the bottom level of the action-centered subsystem, implicit reactive routines are learned in neural networks. Reinforcement learning algorithms may be applied

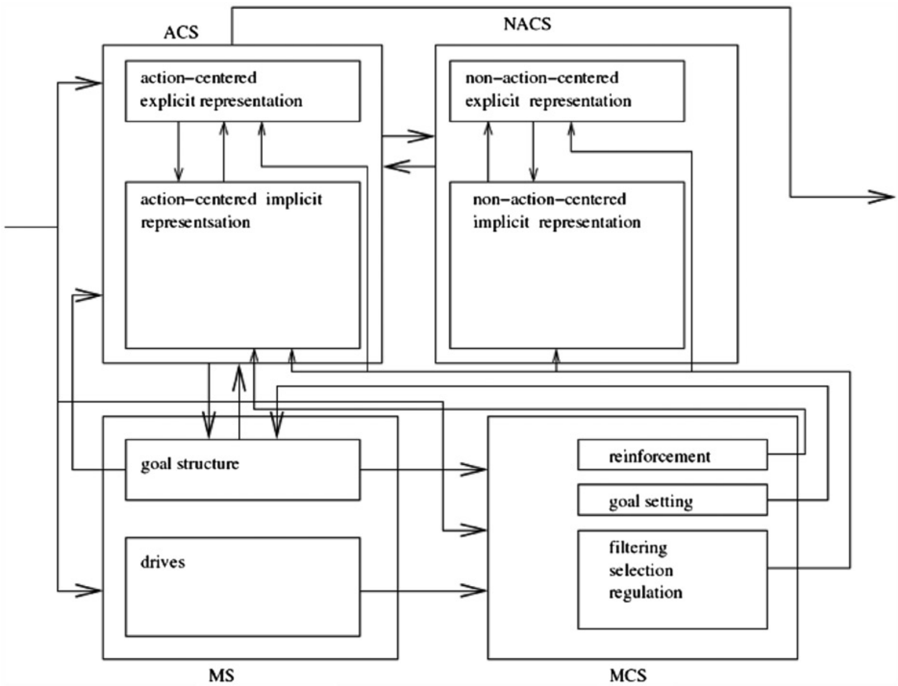

Figure 1. The Structure of CLARION.

(as detailed later). The bottom level thereby develops implicit sequential behaviors (i.e., reactive routines; Sun 2002). The bottom level is modular.

In the top level of the action-centered subsystem, explicit conceptual knowledge is captured in the form of symbolic rules. There are many ways in which explicit symbolic knowledge may be learned, including independent hypothesis-testing learning and bottom-up learning.

As discussed before, implicit knowledge may be acquired through trial and error, and on top of that, explicit symbolic knowledge can be acquired also from on-going experience in the world through the mediation of implicit knowledge; hence *bottom-up* learning (Karmiloff-Smith, 1986; Sun et al., 2001). The basic process of bottom-up learning is as follows: if an action implicitly decided by the bottom level is successful, then the agent extracts an explicit symbolic rule that corresponds to the decision made by the bottom level and adds the rule to the top level. Then, in subsequent interaction with the world, the agent verifies the extracted rule by considering the outcome of applying the rule: if the outcome is not successful, then the rule should be made more specific; if the outcome is successful, the agent may try to generalize the rule to make it more universal. The operations involved may be specified precisely based on statistical measures (more details later).

Although CLARION can learn even when no a priori or externally provided knowledge is available, it can make use of it when such knowledge is available. To deal with such instructed learning, externally provided knowledge in the forms of explicit symbolic representations can (1) be combined with existing symbolic representations at the top level (i.e., internalization; but see Carey, 2011), and (2) be assimilated into implicit reactive routines at the bottom level (i.e., assimilation). This process is referred to as "top-down learning" (more details later; see also Dreyfus & Dreyfus, 1987; Vygotsky, 1962).

3.3.2. The non-action-centered subsystem

The non-action-centered subsystem (the NACS) is for representing declarative (i.e., general) knowledge about the world, and for performing declarative memory retrievals and inferences. The NACS is under the control of the ACS (through its actions).

At the bottom level, "associative memory" networks encode implicit non-action-centered knowledge using subsymbolic representations. Backpropagation (i.e., "Multi-Layer Perceptron" or MLP) networks or Hopfield networks may be used to establish implicit associations. On the other hand, at the top level of the NACS, explicit non-action-centered knowledge is encoded using symbolic representations.

In this subsystem, as in the ACS, chunks (representing concepts) are specified through dimensional values. A node is set up in the top level to represent a chunk. The chunk node connects to its corresponding microfeatures (i.e., dimensional values) represented as individual nodes in the bottom level (constituting a prototype model of concepts). Additionally, links between chunk nodes at the top level encode explicit associations between pairs of chunks, known as associative rules. Explicit associative rules may be learned in a variety of ways (Sun, 2003).

The distinction between semantic memory and episodic memory within the NACS should be mentioned here (this distinction has been well known in psychology; see, e.g., Tulving, 1985). Semantic knowledge (in semantic memory of the NACS) is not tied to specific experiences (although it may be generated as a result of past experiences; see the description later). In contrast, episodic knowledge (in episodic memory of the NACS) is directly tied to specific past experiences, with specific time and other episodic information included as part of the knowledge encoding (more later).

Top-Down or bottom-up learning may take place in the NACS as in the ACS (details later).

3.3.3. The motivational and the meta-cognitive subsystem

The motivational subsystem (the MS) is essentially about "drives" and their interactions (Toates, 1986), which lead to goals and actions. It is concerned with why an agent does what it does. Simply saying that an agent chooses actions to maximize gains or payoffs leaves open the question of what determines these things. Through the motivational subsystem, CLARION addresses the *intrinsic* needs and motives of an agent formed through evolution (in particular, in relation to its survival and its functioning in the social and the physical world), as discussed in the enactive cognition tradition (e.g., Di Paolo, 2005; Froese & Ziemke, 2009), and as explored in social psychology.

Dual motivational representations are in place in CLARION, based on available psychological evidence (Sun, 2009). The explicit goals (e.g., "finding food") of an agent (which are tied to the working of the ACS) may be generated based on internal drive activations (e.g., "being hungry").

While there are low-level drives concerning physiological needs (such as hunger, thirst, and so on), there are also higher-level drives. Some of them are primary, in the sense of being "hard-wired" (such as seeking social status, following social code, fairness, and so on).

The meta-cognitive subsystem (the MCS) is closely tied to the MS. The MCS monitors, controls, and regulates cognitive processes for the sake of performance. Control and regulation may be in the forms of setting goals for the ACS, setting essential parameters, interrupting on-going processes, and so on. Control and regulation can also be carried out through setting reinforcement functions for reinforcement learning within the ACS. All of the above can be done on the basis of drives and goals from the MS. The MCS is also made up of two levels: the top level (explicit) and the bottom level (implicit).

4. Autonomous Symbol Generation in CLARION

Turning specifically to symbol generation within CLARION, there are a number of possibilities:

> Implicit-to-Explicit explication of procedural knowledge: extraction of explicit symbolic action-centered knowledge in the action-centered subsystem of CLARION.
>
> Procedural-to-Declarative transformation: transfer of extracted explicit symbolic action-centered knowledge into explicit symbolic non-action-centered knowledge in the non-action-centered subsystem of CLARION.
>
> Implicit-to-Explicit explication of declarative knowledge: extraction of explicit symbolic non-action-centered knowledge in the non-action-centered subsystem of CLARION.

Let us look into some technical details below.

4.1. Extraction of Symbolic Representations in the ACS

First, some basics of rule representation and subsymbolic learning are explained below, and then the algorithms for bottom-up and top-down learning are described.

4.1.1. Symbolic representations in the ACS

At the top level of the ACS, rules are usually in the following form: "CURRENT-STATE-CONDITION → ACTION." The left-hand side of a rule is a conjunction of individual elements, each of which refers to a dimension of state *x*. Each element specifies a value range. The right-hand side of a rule is an action recommendation, which also consists of a set of dimensional values concerning actions.[1]

In the condition of a rule, all of the specified dimensional values together define a chunk. A chunk is represented as a unitary chunk node at the top level, whereas each of its dimensional values is represented as a (separate) microfeature node at the bottom level. At the top level of the ACS, an action, in the right-hand side of a rule, also constitutes a chunk, which is represented as a chunk node at the top level, the same way as the condition of a rule, while each of its dimensional values is represented as a separate (microfeature) node at the bottom level.

Therefore, each chunk node at the top level is connected to all the specified dimensional values at the bottom level. Dimensional values in the bottom level serve as features or microfeatures of a chunk, and the chunk node at the top level in turn serves to identify and label this set of dimensional values as a whole (thereby constituting a prototype model of concepts).

4.1.2. Subsymbolic reinforcement learning in the ACS

In the bottom level of the ACS, the outputs of the neural networks are Q values. A Q value is an evaluation of the "quality" of an action in a given state: $Q(x, a)$ indicates how desirable action a is in state x, which includes the sensory input and the goal. An agent can choose an action based on Q values in state x.

To acquire Q values, the Q-learning algorithm may be applied (Watkins, 1989). Reactive routines developed in the bottom level of the ACS through reinforcement learning can exhibit sequential behavior without explicit (symbolic) planning (Suchman, 1987).

To implement Q-learning, a four-layered backpropagation (MLP) network may be used. The network is internally subsymbolic. The output of the third layer indicates the Q values (each represented by an individual node), and the fourth layer determines probabilistically the action to be performed based on a stochastic distribution of Q values.

4.1.3. Bottom-up rule learning and generation of symbolic representations within the ACS

In the action-centered subsystem, once the bottom-level reactive routines have been acquired to some extent, an agent may learn explicit rules at the top level using information from the bottom level, which is a bottom-up learning process. This has been referred to as the "Rule-Extraction-Refinement" (RER) algorithm (Sun et al., 2001). This process constitutes the essential part of the autonomous generation of symbolic representations from subsymbolic representations.

The basic idea of this algorithm has been explained earlier. Specifically, the agent does the following at each step:

1. Update the rule statistics.
2. Check the current criterion for rule extraction, generalization, and specialization:

2.1. If the result is successful according to the current rule extraction criterion, and there is no rule matching the current state and action, then perform extraction of a new rule: "CONDITION → ACTION." Add the extracted rule to the top level of the ACS.
2.2. If the result is unsuccessful according to the current specialization criterion, revise all the rules matching the current state and action through specialization:
   2.2.1. Remove the rules from the top level.
   2.2.2. Add the revised (specialized) versions of these rules into the top level.
2.3. If the result is successful according to the current generalization criterion, then generalize the rules matching the current state and action through generalization:
   2.3.1. Remove these rules from the top level.
   2.3.2. Add the generalized versions of these rules to the top level.

One may find psychological arguments in favor of this kind of algorithm in, for example, Bruner et al. (1956) and Sun (2002).

Let us discuss briefly the details of the operations. First of all, at each step, the positive and negative match counts (i.e, $PM_a(C)$ and $NM_a(C)$), are updated (in step 1 of the above algorithm) for each rule condition and each of its minor variations, with regard to the action just performed. That is, $PM_a(C)$ (i.e., Positive Match for C) equals the number of times that a state matches condition C, action a is performed, and the result is positive; $NM_a(C)$ (i.e., Negative Match for C) equals the number of times that a state matches condition C, action a is performed, and the result is negative. Positivity or negativity may be determined based on a certain positivity criterion, which depends on task circumstances (Sun et al., 2001).

Based on these statistics, the information gain measure $IG(A, B)$ may be calculated. Essentially, the measure compares the percentages of positive matches under different conditions A and B. If A can improve the percentage to a certain degree over B, then A is considered better than B. In the algorithm above, if a rule is better compared with its corresponding "match-all rule" (i.e., the rule with the same action but with the condition that matches all possible input states), then the rule is considered successful. For example, if $IG(C, all) > threshold$, then perform generalization, replacing C with C' (a minimum generalization of C that maximizes $IG(C', C)$).[2]

It is important to relate the process above to the issues surrounding symbol generation discussed earlier. In the ACS, a concept may be learned and a symbolic representation of the concept may be established as a result of extracting an action rule in the process of (and for the sake of) accomplishing a particular task. When the condition of an action rule is established, a localist (symbolic, unitary) encoding of the condition is also set up (in the form of a chunk node), and a new symbol is thereby formed. Specifically, when a rule is extracted, a separate node (a chunk node) is set up in the top level of the ACS to represent the condition of the rule as a whole, which connects to its microfeatures (dimensional values) represented individually in the bottom level of the ACS. Together they form a chunk, which is a prototype representation of a concept.

For example, a symbolic chunk node may be set up in the top level of the ACS to represent the following chunk: ((TEMP, WARM) (RAINFALL, HEAVY)). The chunk node is linked to the bottom-level nodes representing (temp, warm) and (rainfall, heavy). This chunk (and the corresponding chunk node at the top level) may be acquired in the process of extracting an action rule in the top level of the ACS, such as "(TEMP, WARM) (RAINFALL, HEAVY) → (ACTION, STAY-UNDER-TREES)," which in turn happens in the process of accomplishing a particular task.

Explicit symbolic knowledge acquired in this way is concerned with existentially significant aspects of the world in relation to the agent involved, to its goals and needs. In other words, symbolic concepts and rules learned in the top level of the ACS are concerned with those aspects of the world that have significant bearings on an agent in its survival and in its interaction with the world. They are not necessarily "objective" classifications, but the result of the interaction of the agent with its world, manifesting the regularities encountered in such interactions (Merleau-Ponty, 1963) and reflecting the intrinsic needs of the agent (Johnson, 1987; Varala et al., 1991; on the basis of the MS of CLARION). They are also action-oriented, concerned specifically with helping to decide on what to do in given situations. Therefore, they are task-oriented, action-oriented, goal/need-oriented, and grounded in the interaction between an agent and its world (Heidegger, 1927/1962; Merleau-Ponty, 1963). In this way, the agent projects its own perspectives and needs onto the world, and in the process brings forth the meanings of the situations and the meanings of the symbolic representations involved (Johnson, 1987).

Sun (2002) provides computational analyses of concepts formed (symbols generated) by agents in the context of learning specific tasks within CLARION. It was found that the concepts (symbols) that had been formed were indeed concerned with those aspects of the environment that were important to the agent's goals and tasks at hand, serving the purpose of facilitating the agent's action decisions in accomplishing a task. Specifically, in the contexts that were analyzed, those concepts (symbols) help to identify certain combinations of environmental features (dimensional values in the bottom level of CLARION) that are significant for the agent's action decision making, thus helping the agent to make proper action decisions under those circumstances, thereby facilitating the accomplishment of its goals.

4.1.4. Top-down learning and assimilation within the ACS

As mentioned before, top-down learning is the process by which externally given explicit knowledge is incorporated into the top level and assimilated into the bottom level. First, externally given explicit knowledge may be expressed as rules and chunks at the top level (in the forms as discussed before). Assimilation into the bottom level can then be done, either through gradual practice (with Q-learning in the bottom level) guided by the top-level knowledge, or by using supervised learning in which the top level serves as the teacher.

Assimilation through gradual practice while guided by the top-level knowledge is always (automatically) performed. With explicit knowledge (in the form of explicit rules) in place at the top level, the bottom level learns at every step under the "guidance" of the rules. Initially, the agent relies mostly on the explicit rules at the top level for its action decision making, and meanwhile learns implicit knowledge at the bottom level through "observing" the actions directed by the top-level rules (with exactly the same reinforcement learning mechanism at the bottom level as described before). But gradually, when more and more implicit knowledge is acquired by the bottom level, the agent relies more and more on the bottom level (given that the inter-level integration is adaptable). Hence, top-down learning takes place (Sun, 2002).

Through such assimilation, explicit symbolic knowledge becomes implicit knowledge (in particular, procedural, embodied skills in the ACS), and thus becomes more effective (Dreyfus & Dreyfus, 1987). Although not autonomous generation of symbolic representations, this process offers a possibility for the grounding of (previously generated, externally given) symbolic representations. Externally given symbolic representations may be grounded in low-level, implicit, subsymbolic representations through the ongoing activities of the agent interacting with the world. That is, they may be assimilated into the low-level implicit representations and enmeshed in the on-going interaction between the agent and the world (which is the overarching context within which assimilations take place). Therefore, they can be used in ways that enhance the survival and functioning of the agent in the world (otherwise, they will not be useful and will be forgotten within CLARION).

4.2. Transfer of Symbolic Knowledge from the ACS to the NACS

Let us turn to the NACS. In the NACS, chunks at the top level (symbols) are formed from a variety of sources, including the following: (1) each state experienced as a whole (as observed by the ACS) is encoded as a chunk in the NACS; (2) so is each action chosen by the ACS; (3) so is each perception-action step experienced as a whole.

Within the NACS, these types of chunks are part of semantic memory. They are experience-based, because they are created due to experiences and they reflect such experiences. In a sense, they are transferred from the ACS to the NACS, or in other words transferred from the procedural memory to the declarative, semantic memory. Therefore, they are also task-oriented, and concerned with the agent-world interaction.

Chunks may also be transferred to the NACS when they are extracted as a result of bottom-up rule learning (using the RER algorithm) in the ACS. As mentioned before, in the ACS a chunk may be learned as a result of extracting an action rule: when the condition of an action rule is established, a localist encoding of that condition is also established at the top level of the ACS. That is, a new chunk (with a new symbol) is formed in the ACS, which connects to its microfeatures represented in the bottom level of the ACS. At the same time as the bottom-up rule learning in the ACS, a corresponding semantic chunk is set up in the NACS, and its chunk node is linked to the similar bottom-level representations in the NACS. This is another instance of the transfer of procedural knowledge to declarative knowledge (from the ACS to the NACS).

Symbols (denoting concepts) that have been formed in the ways above, despite the fact that they reside in the NACS, are context-dependent and goal/task-oriented, because they are formed in relation to tasks and goals at hand and for the purpose of exploiting environmental regularities encountered in dealing with a task (by the ACS). Thus, even in the NACS, CLARION emphasizes the functional role of symbols/concepts and the importance of goal/need in forming symbols and concepts: a symbol (or a concept) is formed as part of an action rule (or simply an action in a situation) in the ACS, which is learned to accomplish a task in a particular environment. As a result, acquired symbols/concepts are functional, even when they are transferred to the NACS. Task and goal/need contexts help an agent to determine which set of microfeatures in the environment need to be attended to together. Symbolic knowledge acquired in this way is concerned with existentially and ecologically significant aspects of the world: those aspects of the world that have significant bearings on an agent in its interaction with the world and ultimately in its survival (Johnson, 1987; Varela et al., 1991). They are not necessarily "objective" classifications of the world, but the result of the interaction of an agent with its world and the agent's projection.

When two-level dual representations of chunks in semantic memory are established in the NACS, they constitute a prototype model of concepts. Localist chunk nodes at the top level of the NACS serve as the identification of a set of correlated microfeatures at the bottom level, in a bottom-up direction. Chunk nodes at the top level also serve to trigger microfeatures at the bottom level, in a top-down direction, once a relevant symbolic concept is brought into attention (i.e., activated).

As discussed earlier, the distinction between semantic and episodic memory has been well argued for in psychology (e.g., Tulving, 1985). While semantic memory is not tied to specific past experiences (although it may be generated as a result of past experiences, as those types of semantic chunks above), episodic memory is directly tied to specific past experiences in an individuated form (with specific time and other episodic information included as part of the encoding). In CLARION, there are the following types of episodic memory items: (1) each state (at each step) as a whole, as observed by the ACS, is represented as a chunk in the episodic memory, along with a time stamp; (2) each action chosen by the ACS at each step is represented as a chunk in the episodic memory, along with a time stamp; and (3) each step performed by the ACS, as a whole, is represented as a chunk in the episodic memory, along with a time stamp.

In addition, once any entity (a condition or action chunk, an action rule, and so on) is invoked in the ACS, it leads to the creation of a corresponding episodic memory item (with the same internal makeup plus a time stamp). Clearly, the aforementioned episodic items are transferred from the ACS to the NACS (to the episodic memory). Moreover, judging from the descriptions above, both semantic and episodic knowledge (in semantic and episodic memory, respectively) are task-oriented, and result from agent-world interaction.

4.3. Symbolic Knowledge Extraction in the NACS

In the NACS, chunks acquired from sources outside of the NACS (in ways discussed above) are used to encode explicit knowledge extracted from the bottom level of the NACS, in the form of associative rules between chunk nodes.

Specifically, within the NACS, an associative rule may be extracted when an associative mapping is performed in the bottom level of the NACS, in which case associative links are established at the top level of the NACS between the chunk node denoting the cue for the mapping and chunk nodes denoting outcomes from the mapping.

More specifically, a number of chunks are set up and activated: (1) a cue used in associative mapping in the bottom level of the NACS is encoded as a chunk (with a chunk node, a symbol, at the top level of the NACS) and activated, if there is no chunk already there corresponding to the cue (with the same set of dimensional values); (2) through bottom-up activation, the result from the associative mapping at the bottom level of the NACS activates at the top level of the NACS all the currently existing chunks compatible with (i.e., contained within) the result; (3) a separate chunk is set up and activated that corresponds to the result from the bottom level of the NACS as a whole (if such a chunk is not already there); (4) a separate chunk is also set up and activated that corresponds to the cue (all its dimensional values) and the result (all its dimensional values) together as a whole (if such a chunk is not already there).[3]

At the same time, various associative rules corresponding to the associative mapping happening at the bottom level of the NACS are also set up at the top level of the NACS (if not there already). That is, an associative link is established that connects the chunk node representing the cue with each chunk node compatible with the result of the bottom-level mapping (of a type specified by (2) or (3) above).

In addition, as with the ACS, once any entity (a chunk, an associative rule, and so on) is invoked in the NACS, it leads to the creation of a corresponding episodic memory item (a symbol, with the same internal makeup plus a time stamp) in the episodic memory of the NACS.

Similar to concept learning in the ACS as discussed before, chunks (denoting concepts) and associative rules formed in the ways specified above in the NACS are also task-oriented (to an extent), because they are formed in relation to tasks at hand when the NACS capacities are invoked by the ACS in its dealing with specific tasks.

5. Examples of Symbol Generations

Below are some thought experiments utilizing these forms of bottom-up extraction and top-down assimilation. They involve the interaction of these different ways of symbol generation and grounding during an agent's interaction with the world (sociocultural or physical).

5.1. Learning about KNIFE*: An Example*

Imagine learning the concept KNIFE (and related knowledge). Unaware of the danger of a knife, a young child may approach the sharp edge of a knife, which causes pain. Recoiling from the object, the child quickly registers a rule: this thing is to be avoided. Soon enough, he forgets that rule (because he has so many other things to remember). So the experience re-occurs a few of times under similar or different circumstances. The repeated experiences lead the child to develop an intuition and a reactive routine (on the basis of innate perceptual and conceptual primitives; Carey, 2011): stay away from sharp edges.

On the other hand, parental input provides the child with a verbal label for the object that can cause pain (Vygotsky, 1962): "knife" (while pointing to the object). The label initially may be closely associated with the visual image of a particular knife and the pain that it caused. But gradually, it is revised and generalized in accordance with experiences (Vygotsky, 1962): 'knife' could be in various shapes (but always having a sharp edge), could be in various sizes, has a handle, and so on (including various visual, tactile, proprioceptive, and haptic information; Barsalou, 1999).

Implicit intuition about KNIFE and implicit reactive routines associated with it thereby develop gradually. Then, through bottom-up explication, explicit rules concerning what a knife is and how one should act in relation to it are also developed (Karmiloff-Smith, 1986; Sun et al., 2001).

The establishment of explicit symbolic concepts and rules concerning KNIFE by the child leads to various further knowledge, beliefs, and memories associated with it, and consequently various kinds of reasoning that can be performed about it (Piaget, 1971).

Later it may occur to the child that if he needs to slice a tomato, knives may be useful. Going further, it may occur to him that if he needs to kill an animal, knives may also be used. From that point on, the child may find many uses for it and develop much knowledge about it (Helie & Sun, 2010).

A similar description of learning may apply to a wide range of concepts, from stone-age tools to modern technical devices.

5.2. Learning about KNIFE *within CLARION*

Let us examine how CLARION carries out the afore-described processes.

5.2.1. Symbol and rule extraction in the ACS

First of all, in the ACS, implicit reactive routines are developed in the bottom level through reinforcement learning with repeated, trial-and-error experiences, which form the basis for actions, for example, in the presence of knives.

Based on such implicit reactive routines, explicit concepts and explicit action rules arise within the ACS, for example, through the RER algorithm, that is, through extracting and refining explicit action rules at the top level based on the information from the bottom level.

Using the RER algorithm as described before, the following explicit action rules may be created within the ACS: "SHARP EDGE, SHINING METAL SURFACE, HANDLE → DO NOT TOUCH." Such rules are concerned with existentially significant aspects of the world.

Note that rules are extracted through the various operations involving extraction, generalization, and specialization (as described before, as part of the RER algorithm). For example, initially, the following rule might have been extracted: "LONG SHARP EDGE, SHINING METAL SURFACE, WOODEN HANDLE → DO NOT TOUCH." Then, through generalization, it became: "LONG SHARP EDGE, SHINING METAL SURFACE, HANDLE → DO NOT TOUCH." Further generalization and specialization led to: "SHARP EDGE, SHINING METAL SURFACE, HANDLE → DO NOT TOUCH."

As a result of rule extraction, symbols (i.e., chunk nodes, indicating chunks representing concepts) have been formed in the process of extracting and refining the rules, for example, the concept KNIFE here (which consists of sharp edge, shining metal surface, handle, etc.). As discussed (more theoretically) before, such symbols are meaningful, because (1) they are linked to the bottom-level subsymbolic representations, and (2) they were created in the process of accomplishing an existentially relevant task (such as dealing with the world around), and therefore they are significant to the agent and its dealing with the world.

5.2.2. Symbol transfer into the NACS

Those symbols (indicating concepts, with chunk node representation) extracted within the ACS may also be useful outside of the ACS. For example, they enable the agent to reason within the NACS, explicitly or implicitly, about relevant situations. Due to the existential relevance of these symbolic representations, their presence within the NACS renders the working of the NACS also existentially relevant—they provide relevance to reasoning and other functionalities performed within the NACS. Therefore, symbolic processes within the NACS are also grounded in the agent's subsymbolic activities and the agent-world interaction.

For example, a symbol (a chunk node denoting a concept) transferred from the ACS may represent KNIFE. It is then used in the NACS for constructing explicit non-action-centered knowledge (e.g., associative rules). For example, an associative rule may be as follows: "KNIFE, HOSTILE PERSON → POTENTIALLY VIOLENT SITUATION," or "KNIFE, HOSTILE PERSON → DANGEROUS SITUATION" (provided that other symbols/chunks involved, such as HOSTILE PERSON, POTENTIALLY VIOLENT SITUATION, and DANGEROUS SITUATION have already been established within the NACS). Such symbols and rules facilitate reasoning about a situation, particularly in an explicit, deliberative manner (that is, at the top level, but also in an implicit, intuitive manner at the bottom level of the NACS). For instance, the associative rule above may be used, along with other possible rules and concepts, for explicit reasoning about various options in a dangerous standoff.

5.2.3. Symbol extraction within the NACS

Symbols (denoting concepts) may be extracted within the NACS itself. For instance, in the afore-described domain, a concept (hence a symbol) may be formed as a result of implicit associative mapping in the bottom level of the NACS (as one of the outcomes from associative mapping specified earlier): POTENTIALLY VIOLENT SITUATION (which involves microfeature representations). At the same time, associative rules, such as "KNIFE, HOSTILE PERSON → POTENTIALLY VIOLENT SITUATION," may be extracted also from implicit associative mapping performed at the bottom level of the NACS (as described before).

5.2.4. Interaction of external and internal symbols

In the ACS, externally provided information may enable top-down learning (as opposed to bottom-up learning). For example, a parent may instruct the child (while pointing to a knife): "if you see a knife lying around, don't touch it." This instruction may be set up at the top level of the ACS as an action rule, with its condition and conclusion set up as two separate chunks (with chunk nodes at the top level, i.e., with symbols). The chunk nodes at the top level are linked to the bottom-level subsymbolic microfeature representations (from perceptual, motor, and other primitives). The action rule may be further assimilated into the bottom level of the ACS (into its implicit reactive routines). It may also be further refined as described before. Of course, this is not exactly autonomous learning, but it is a common, readily present part of learning and development for humans.

Assimilation of externally provided symbols and symbolic knowledge also occurs within the NACS. For instance, externally provided information (e.g., from a parent) may be: "knives are made of metal." This information may be set up at the top level of the NACS as an associative rule: "KNIFE → METAL-OBJECT." Related chunk nodes (denoting concepts), for example, one representing the condition and the other the conclusion, are also established and connected to the subsymbolic representations at the bottom level of the NACS (which give the symbols their meanings). In this case, while the chunk for KNIFE may have already been established in the ACS and thereafter transferred into the NACS, the chunk for METAL-OBJECT may be new and therefore established as a result of this assimilation process.

In the process, externally provided symbols, extracted symbols, and transferred symbols (from the ACS) interact within the NACS. For instance, an external symbols may be subsumed by an extracted symbol, or vice versa, and thereby the two kinds of symbols are related to each other, and as a result enhance each other (Carey, 2011). As an example, an externally provided symbol may denote a concept of GANG RIVALRY, and this symbol may be subsumed by (i.e., considered a subcategory of) the internally extracted symbol (concept) POTENTIALLY VIOLENT SITUATION (as discussed earlier), which is in turn a subcategory of DANGEROUS SITUATION. For another example, imagine that the child was told that knives are made of metal, while he himself previously extracted the associative rule that knives were hard objects. So he might infer that metal objects are hard objects (or vice versa) within the NACS.

As a result of such interactions, the conceptual system of the agent (with symbolic and subsymbolic representations from both the ACS and the NACS) becomes richer and more complex. There is a chance that the enriched symbolic representations may be more broadly useful, and thus spread culturally and enrich culture-wide representations.

The upshot is that for any individual, not all concepts and symbolic representations that the individual possesses are acquired culturally (or externally). Some representations in the mind of an individual are learned autonomously (on the basis of innate perceptual and conceptual primitives), although such representations interact with culturally prevalent symbols, concepts, and representations as well as being closely related to the individual's social interaction. The danger of downplaying the role of autonomous learning and autonomous generation of symbolic representations is that we may end up mistakenly viewing individuals as robots being entirely programmed by the culture in which they found themselves, and neglecting other, potentially equally important possibilities.

6. Discussion

Symbolic representations established in CLARION are goal/need-oriented, task-oriented, context-dependent, and action-relevant. In other words, they are not arbitrary, but pertinent to the viability, survival, and functioning of the agent in the world, and geared towards the agent interacting with and surviving in the world, reflecting its existential concerns and the corresponding valuations.

Furthermore, given the above, the specific roles of symbolic representations (in the top levels of CLARION, in both the ACS and the NACS) may include the following: (1) as the knowledge extractor; (2) as the interpreter; (3) as the communicator; and (4) as the data compressor/encoder. For instance, psychological experiments show that symbolic labels highlighted the commonalities between objects, facilitated categorization, and even overrode natural perceptual categories. These roles may justify the existence of symbols in cognitive agents, in an evolutionary sense.

Given the various roles and, consequently, the importance of symbolic representation, we may further examine the importance of autonomous, bottom-up generation of symbols on the basis of embodied interaction with the world. Bottom-up generation may be significant in two senses: the ontological sense and the ontogenetic sense. Ontologically, explicit symbolic knowledge needs to be obtained in the first place before it can be imparted to other individuals to enable, for example, top-down learning. Therefore, bottom-up symbolic knowledge extraction, which creates new symbolic knowledge, is more fundamental. Only after bottom-up symbolic knowledge extraction (and other types of learning) create explicit, symbolic, conceptual knowledge, can top-down learning be possible. Practically speaking, given the culturally created systems of schooling, apprenticeship, and other forms of guided or instructional learning, top-down learning is quite prevalent in human society. However, it should be emphasized that bottom-up learning is more fundamental in the sense described above.

Ontogenetically, there appear to be some empirical (psychological) indications that children learn sensory-motor skills (as well as possibly knowledge concerning concepts and so on) implicitly and subsymbolically first, and then acquire explicit symbolic knowledge on that basis (see Karmiloff-Smith, 1986, for summaries of various such empirical indications). Therefore, beyond ontological significance, bottom-up symbolic knowledge extraction is also significant ontogenetically (developmentally) in a psychological sense.

It is useful to point out that implicit reactive routines ("comportment" in the Heideggerian sense) that are relied upon in autonomous symbol generation via bottom-up learning are not the same as the "embodiment" that has been advocated before. Lakoff and Johnson (1980) have put forth a view that cognition is largely determined by the bodily basis of cognitive agents: bodily schemata get abstracted and mapped onto other domains. From that view, an object should not be understood and represented only in terms of its shape, color, or any other static features, but should be approached also in terms of what an agent can do with it. However, they leave open too many possibilities. For example, there are too many different uses we can make of a cup: we can drink from it; we can use it to store water, powder, coins, paper clips, or business cards; we can hold it in our hands; we can put it on top of our heads; and so on. Its uses are almost limitless. How can an agent structure its understanding around so many different possibilities? The key, according to CLARION, is what an agent commonly, routinely, and reflexively does with an object in its everyday life, that is, what we called "comportment" (of an agent) toward an object (being-with-things; Heidegger, 1927/1962), which is made up of implicit reactive routines in CLARION. Such comportment with objects is the basis of how an agent approaches objects, reflecting its needs and goals in its effort to survive and to function in its world (and thus its corresponding perspectives and valuations).

There is some existing work that is more or less relevant to this approach. The existing work includes conceptual grounding in, for example, Barsalou (1999) and Glenberg and Kaschak (2002). These projects, although going in the right direction, do not provide detailed, process-based, mechanistic theories. On the other hand, Steels (2003) and Cangelosi (2010) do provide computational algorithms. However, these projects do not go far enough in meeting the essential desiderata discussed earlier. For example, they often do not make clear enough distinctions between implicit and explicit psychological processes, and do not deal with bottom-up learning as elucidated earlier. Many of them do not take the essential needs and motives of agents into sufficient consideration.

Finally, note that in relation to symbol generation, various detailed psychological simulations have been carried out within CLARION and described in various technical papers (see, e.g., Sun et al., 2001, 2005).

7. Concluding Remarks

This paper has described a number of mechanisms for creating symbolic representations in the course of interacting with the world, on the basis of situated/enactive cognition (hence meaning emerges), within a comprehensive framework for modeling the human mind. The significance of this approach lies in generating symbolic representations through (in the main) autonomous symbolic generation from subsymbolic processes, and demonstrating it through detailed mechanistic, process-based (computational) accounts in a psychologically realistic cognitive architecture (for psychological validity of CLARION, see Helie & Sun, 2010; Sun et al., 2001, 2005).

The philosophical implications of this work lie in the substantiation of a unifying perspective that embraces both situated interaction as well as symbolic representation (on top of, and based upon, situated interaction). This perspective corrects some views of cognition that shun symbolic representations altogether, and attempts to be more philosophically even-handed, more psychologically realistic, and more computationally comprehensive, for the sake of better understanding of the human mind.

Acknowledgements

This research has been supported in part by the ONR grant N00014-08-1-0068. Thanks are due to my many collaborators, past and present.

Notes

[1] Alternatively, the rules can be in the forms of "CURRENT-STATE-CONDITION → ACTION NEW-STATE" or "CURRENT-STATE-CONDITION ACTION → NEW-STATE." The encoding of the alternative forms of rules is similar.

[2] Here we need to address the issue of proliferation of rules and chunks. In CLARION, an encoding probability is specified so that not all such entities may be encoded, and an interval for forgetting (known as the density parameter) is also specified so that dormant entities may be removed. Another issue is proliferation of states resulting from a large number of dimensions, in which case attentional focusing may be necessary, which may be accomplished by the MCS in CLARION. The same approach applies to other representational entities discussed later.

[3] The chunk is formed by merging the dimensional values of the cue with those of the result.

References

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

Bruner, J., Goodnow, J., & Austin, J. (1956). *A study of thinking*. New York: Wiley.

Cangelosi, A. (2010). Grounding language in action and perception: From cognitive agents to humanoid robots. *Physics of Life Reviews*, *7*, 139–151.

Carey, S. (2011). The origin of concepts: A précis. *Behavioral and Brain Sciences*, *34*, 113–167.

Dewey, J. (1958). *Experience and nature*. New York: Dover.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*, 429–452.
Dreyfus, H. (1992). *Being-in-the-World*. Cambridge, MA: MIT Press.
Dreyfus, H. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology*, *20*, 247–268.
Dreyfus, H., & Dreyfus, S. (1987). *Mind over machine: The power of human intuition*. New York: The Free Press.
Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systematic organization of life and mind. *Artificial Intelligence*, *173*, 466–500.
Glenberg, A., & Kaschak, K. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558–565.
Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York: Harper and Row. (Original work published 1927).
Heidegger, M. (1984). *The metaphysical foundations of logic* (M. Heim, Trans.). Bloomington, IN: Indiana University Press. (Original work published 1978).
Helie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, *117*, 994–1024.
Johnson, M. (1987). *The body in the mind: The bodily basis of imagination, reason, and meaning*. Chicago: University of Chicago Press.
Johnson-Laird, P., & Yang, Y. (2008). Mental logic, mental models, and computer simulations of human reasoning. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 339–358). New York: Cambridge University Press.
Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, *23*, 95–147.
Lakoff, G. (1988). Cognitive semantics. In U. Eco, M. Santambrogio, & P. Violi (Eds.), *Meaning and mental representations* (pp. 119–154). Bloomington, IN: Indiana University Press.
Lakoff, G., & Johnson, M. (1980). The metaphoric structure of human conceptual system. *Cognitive Science*, *4*, 193–208.
Masmoudi, S., Dai, D.Y., & Naceur, A. (2012). *Attention, representation, and human performance: Integration of cognition, emotion, and motivation.* New York: Psychology Press.
Merleau-Ponty, M. (1963). *The structure of behavior*. Boston: Beacon Press.
Piaget, J. (1971). *Biology and knowledge*. Edinburgh, UK: Edinburgh University Press.
Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.
Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences*, *7*, 308–312.
Suchman, L. (1987). *Plans and situated actions*. Cambridge: Cambridge University Press.
Sun, R. (1994). *Integrating rules and connectionism for robust commonsense reasoning*. New York: John Wiley and Sons.
Sun, R. (2000). Symbol grounding: A new look at an old idea. *Philosophical Psychology*, *13*, 149–172.
Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Lawrence Erlbaum Associates.
Sun, R. (2003). *A tutorial on CLARION* (Technical report). Troy, NY: Rensselaer Polytechnic Institute, Cognitive Science Department.
Sun, R. (2009). Motivational representations within a computational cognitive architecture. *Cognitive Computation*, *1*, 91–103.
Sun, R. (2012). Memory systems within a cognitive architecture. *New Ideas in Psychology*, *30*, 227–240.
Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*, 203–244.
Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, *112*, 159–192.
Toates, F. (1986). *Motivational systems*. Cambridge: Cambridge University Press.

Tulving, E. (1985). How many memory systems are there?. *American Psychologist*, *40*, 385–398.
Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.
Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.
Watkins, C. (1989). *Learning with delayed rewards*. Unpublished doctoral dissertation, Cambridge University.
Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., & Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, *291*, 599–600.