Symbol Grounding: A New Look At An Old Idea

Ron Sun
CECS Department
University of Missouri - Columbia
Columbia, MO 65211, USA
rsun@cecs.missouri.edu

NEC Research Institute
4 Independence Way
Princeton, NJ 08540

Appeared in: *Philosophical Psychology*, Vol.13, No.2, pp.149-172. 2000.

Abstract

Symbols should be grounded, as has been argued before. But we insist that they should be grounded not only in subsymbolic activities, but also in the interaction between the agent and the world. The point is that concepts are not formed in isolation (from the world), in abstraction, or "objectively". They are formed in relation to the experience of agents, through their perceptual/motor apparatuses, in their world and linked to their goals and actions.

In this paper, we will take a detailed look at this relatively old issue, using a new perspective, aided by our new work of computational cognitive model development. To further our understanding, we also go back in time to link up with earlier philosophical theories related to this issue. The result is an account that extends from computational mechanisms to philosophical abstractions.

1 Introduction

Symbols and symbol manipulation have been central to cognitive science (Newell and Simon 1976, Minsky 1983, Fodor 1975, Fodor and Pylyshyn 1988). But the disembodied nature of traditional symbolic systems has been troubling many cognitive scientists (Searle 1980, Churchland 1986, Winograd and Flores 1987, Dreyfus and Dreyfus 1987, Agre 1988, Waltz 1990, Wertsch 1991, Bickhard 1993, Varela et al 1993, Sun 1994, Freeman 1995). One remedy that has been proposed is symbol grounding, that is, connecting symbols to lower-level sensory-motor procedures and thus rooting the abstract in the concrete (Harnad 1990, Barsalou 1999).

In this paper, we will take a detailed look at this old issue. Aided by our new work on computational (in a broad sense) cognitive modeling, which we have previously published extensively on (Sun 1997, Sun and Peterson 1998 a, b, Sun et al 1998 a, b), we hope to further our understanding of this issue. We will do so by utilizing a concrete example of a model that offers a new perspective on matters related to this issue and beyond. The new ideas to be offered from this model include the hypothesis of dual processes (mediated and unmediated interaction), bottom-up learning (a two-stage process of learning symbolic representation), and symbol grounding in direct "comportment". Armed with our new perspective, we will also go back in time, to link up with traditional philosophical accounts relevant to this issue, such as Heidegger (1927a).

My main points can be summarized as follows. Symbols should be grounded, as has been argued for many years. But we insist that they should be grounded not only in subsymbolic activities (subsymbolic "representation"), but also in the direct interaction between the agent and the world. The point is that concepts, which symbols represent, are not formed in isolation (from the world), in abstraction, or "objectively". They are formed in relation to the life-world of agents, through the perceptual/motor apparatuses of agents, linked to their goals, needs, and actions. This view is argued on the basis of Heideggerian philosophy, which emphasizes the primacy of direct, unmediated interaction between agents and the world. Symbolic representation and concepts are derived from such direct interaction. Precisely in this sense, can symbols really be grounded. We will show how this can be achieved computationally.

The remainder of this paper is a detailed account that extends from computational (in a broad sense) mechanisms to philosophical abstractions. In section 2, I will review the traditional notions of symbols, representation, and so on, and identify their important properties. In section 3, I will offer a theoretical perspective on these issues that remedies shortcomings of traditional approaches, drawing ideas from Heideggerian philosophy. In section 4, the mechanistic (computational) underpinning of this perspective will be analyzed and viable implementations suggested. In section 5, the framework thus far outlined will be analyzed in light of the issues raised in the first section. In section 6, further discussions will complete the paper.

2 Symbols and Representation: Some Background

In this section, I will review some background notions needed in our discussion, including the notions of symbols and representation. I will try to clarify a few possible confusions and lay the foundation for later exposition of our new perspective (and our new models). In so doing, I will identify a few relevant properties of each of these notions, all of which I will utilize later on.

2.1 Symbols

Symbols have been a mainstay of cognitive science (ever since its early incarnations as information processing psychology and AI in the 1950's). The idea is based on the notion of computation as commonly understood in the early days (so in some sense, it is based on a simplistic and narrowly conceived notion of computation). Computation consists of input, output, retrieval, storage, and manipulation of symbols (such as inside a von Neuman digital computer); a sequence of specified steps accomplishes a computational task, such as numerical calculation. Cognition is understood in much the same way, as consisting of input/output, storage/retrieval, and symbol manipulation. The most important part of this is, of course, the manipulation of symbols that changes symbol structures that stand for mental states. Because of the use of this computer metaphor, cognition is perceived as a sequence of explicit and clear-cut steps involving nothing but symbols.

Later, the *physical symbol system hypothesis* introduced by Newell and Simon (1976) clearly articulated this "vision" and called for a concentrated research program along the symbol manipulation line. They claimed that "symbols lie at the root of intelligent action". They defined, as the fundamental building block of the science of the mind, *physical symbol systems*:

> A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (symbol structure). Thus a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another).

They further claimed that symbols can designate arbitrarily: "a symbol may be used to designate any expression whatsoever"; "it is not prescribed a priori what expressions it can designate." "There exist processes for creating any expression and for modifying any expression in arbitrary ways". Based on that, they concluded: "A physical symbol system has the necessary and sufficient means for general intelligent action", which is the famed physical symbol system hypothesis (where we take "general intelligent action" to mean the full range of human intelligent behavior; Newell and Simon 1976). Clearly, a physical symbol system is an abstracted view of a digital computer (that is, it is an instance of a Turing machine, which is hypothesized in turn by Turing to be able to capture any "computational process"; see Turing 1950). Now the loop is almost complete: If you believe in some kind of universality of "computation" (especially in cognition), and if you believe in Turing's hypothesis (the universality of Turing machines), then you are naturally inclined to believe in the physical symbol system hypothesis (because of the fact that physical symbol systems, as defined by Newell and Simon, are Turing equivalent).

The physical symbol system hypothesis has spawned (and was used to justify) enormous research effort in traditional AI and cognitive science. This approach (i.e., classical cognitivism) typically uses discrete symbols as primitives, and performs symbol manipulation in a sequential and deliberative manner. Although this view came to dominance in AI and cognitive science for a while, it has been steadily receiving criticisms from various sources (for example, from Dreyfus 1972, Dreyfus and Dreyfus 1987, Winograd and Flores 1987, Bickhard 1993, and Searle 1980). They focused on the disembodied abstractness of this approach. In response to such criticisms, Vera and Simon (1993) presented a modified version of physical symbol systems as follows: "A physical symbol system is built from a set of elements, called symbols, which may be formed into symbolic structures by means of a set of relations." "A physical symbol system interact with its environment in two ways: (1) it receives sensory stimuli from the environment that it converts into symbol structures in memory; and (2)

it acts upon the environment in ways determined by symbol structures that it produces." It "has a set of information processes that form symbol structures as a function of sensory stimuli" as well as "produce symbol structures that cause motor actions". Clearly, in this new version, they tried to put more spins on sensory-motor connections, through which (they hope) symbols can be put in contact with the world. Can any of these two versions of the physical symbol system hypothesis justify the claim to universality? Can the new version sustain better the claim?

In this regard, let us look into a basic question: What are symbols after all? "Symbols are patterns", according to the new version of Vera and Simon (1993), in the sense that "pairs of them can be compared and pronounced alike or different." Patterns are symbols when "they can designate or denote". Then, the question becomes: What is the difference between symbols and, say, pictures? According to the old version of the physical symbol system hypothesis, their answer would be that symbols, different from non-symbols such as pictures, can designate arbitrarily; according to the new version, however, their answer is likely to be that there is really no difference. The old version seems overly restrictive: Why do we restrict ourselves to a particular type of pattern and forego the others? How can we believe that such a restricted type of pattern is necessary and sufficient for cognition? The new version seems overly liberal: If anything is a symbol, then of course cognition can be modeled by symbol manipulation; it becomes a tautology and thus trivially true.

In order to pinpoint precisely what a symbol is, we should abstract its essential characteristics. Those characteristics that are at issue here are the following: (1) arbitrariness: whether a pattern (or a sign) has an intrinsic meaning or not; and (2) syntacticity: whether a set of patterns or signs can be arbitrarily and systematically combined in a language-like manner (i.e., whether they have the compositionality and systematicity found in human languages). These two characteristics constitute necessary conditions for a pattern (or a sign) to be a symbol. [1] It is important to emphasize the distinction between the two different notions: signs (generic patterns) and symbols, which Peirce made clear a century ago (see Peirce 1955). To quote from Peirce (1955):

> A sign is either an icon, an index, or a symbol. An icon is a sign which would possess the character which renders it significant, even though its object has no existence; such as a lead-pencil streak as representing a geometrical line. An index is a sign which would, at once, lose the character which makes it a sign if its object were removed, but would not lose that character if there were no interpretant. Such, for instance, is a piece of mould with a bullet-hole in it as sign of a shot..... A symbol is a sign which would lose the character which renders it a sign if there were no interpretant. Such is any utterance of speech which signifies what it does only by virtue of its being understood to have that signification.

Because of the nonsensical consequence of the new version of "symbols", we will have to rely on the old version (Newell and Simon 1976).

Whether symbols are necessary and sufficient for accounting for cognition is not a settled matter. However, almost nobody disputes that some form of symbols, in connectionist, classical, or some other ways, is needed for accounting for high-level cognition, for example, language and reasoning. Even radical connectionists accept that symbols may emerge from dynamic interaction of elements of a

[1] This had been the way symbols were commonly conceived in the cognitive science community (Collins and Smith 1988, Posner 1989), until when the use of symbols and the physical symbol system hypothesis begun to come under attack from connectionists (Bechtel and Abrahamsen 1991, Chalmers 1990, Clark 1993, Sun and Bookman 1994) and situated cognition advocates (Lave 1988, Suchman 1987, Agre 1988, Brooks 1991, Brooks and Stein 1994). Then, all of a sudden, the definition of symbols was drastically altered and enlarged (which renders the notion useless).

network (or a dynamic system in general). I do not make any claim here as to what form symbols should be in, but that there should be some for obvious reasons (Sun 1994). [2]

2.2 Representation

Classical cognitivism believes that an agent has an internal copy, an explicit model of some sort, of the external world. In that model, there are internal states that encode external states in the world (Fodor and Pylyshyn 1988). Cognition is accomplished by developing, maintaining, and modifying these internal states, aka, representation (which is the basic tenet of representationalism). According to the analysis by Peirce (1955), a representational system consists of, or can be analyzed into, representational media, representational entities (i.e., what constitutes a representation), representational semantics (or references, i.e., what is being represented).

To understand characteristics of (explicit) representation (as in traditional cognitive science; Fodor 1975, Collins and Smith 1988), we can identify the following syntactic properties: First of all, such representation is explicit. This is because, without this requirement of explicitness, everything is a representation and thus the tenet of representationalism becomes meaningless. For example, a tennis ball has a representation of forces hitting it and trajectories it flies through, since it can respond to forces and fly through space; a car has a representation of roads and driving movements of its driver, since it can follow the driver's direction and stay on the road (Port and van Gelder 1995). Thus, the representation thesis (representationalism) becomes trivially true in this way. Second, explicit representation is punctate: It consists of clearly delineatable items. Third, it is also elaborate: It contains much detail, even though it may not be a complete model. [3] Forth, explicit representation, as has been traditionally used, is symbolic and compositional. Note that, although explicit representation *need not* be symbolic in the full sense of the term (exceptions include imagery, analogue, and so on), almost all the existing representational models and systems in traditional cognitive science were symbolic (see Anderson 1983, Minsky 1983, Klahr et al 1987 and Davis 1990 for accounts of traditional systems). [4] In addition, explicit representation is semantically specifiable: Each particular representation can have a specific meaning as (arbitrarily) assigned, and meanings are compositional too.

There are many possibilities in terms of manipulating (explicit) representation. Symbol manipulation is the prime candidate for such a task. The compositionality of the syntax and semantics of explicit representation makes it easy to construct a computational procedure to manipulate representation and keep track of references and meanings. There are also other possibilities; for example, connectionist models or neural networks can be somewhat suitable for manipulating representation (see Chalmers 1989, Bechtel and Abrahamsen 1991, Clark 1993, Sun and Bookman 1994).

[2] Only for reason of convenience, I used localist encoding of symbols in my model (Sun 1994, 1997), which will be discussed later on.

[3] A particular version of representationalism advanced by Fodor (1980) is that mental states are represented by propositional attitudes, which include propositions and the agent's relations to them, described in sentential (linguistic) forms.

[4] Note that if it is symbolic, it must be compositional; if it is not, it can still be, and often is, compositional, as we see in existing models.

2.3 Intentionality

One of the most important questions concerning representation is the following: In virtue of what does representation have or acquire its meanings or signification? How does it come to represent what it purports to represent? [5] As we discussed earlier, it is doubtful that an arbitrarily constructed symbolic representation can suffice, by itself, to account for cognition, when only complex relations among symbols are involved (Newell and Simon 1976). Clearly, meanings can hardly lie solely in symbols and their interrelations. How is it possible to move from the notion of meaning as arbitrary designation to a notion of "intrinsic" meaning? In other words, how do we relate representation to whatever is out there in the world, in an intrinsic way? No argument more clearly demonstrates this point than Searle's Chinese Room argument (Searle 1980). The issue brought to light by this argument is *intentionality*. That is, our mental states and mental representation are *about* something. Mere symbols, as pointed out by Searle (1980), have no intentional content and thus cannot capture cognition adequately.

Cognitive science, especially AI, has been grappling with the issue of intentionality ever since the publication of Searle's (1980) argument. [6] Where may an answer to these questions lie? I would venture to suggest a few possible places that we may go to look for answers. These places include: (1) The existential experience of cognitive agents, individually or collectively, especially (but not exclusively) their everyday activities; (2) their functional pre-dispositions in such activities (which are acquired through evolutionary and other processes), including the biological substrates that embody their representational structures, functional capacities, and behavioral propensities. [7] Symbol grounding in the sense of linking symbols (symbolic representation) to lower-level (subsymbolic) processes (Harnad 1990) provides a partial answer to the intentionality question. But it does not fully answer the question. This is because the issue of *how* grounded symbols, and associated subsymbolic processes, acquire their intentional content remains. I would like to argue that, instead of being narrowly and technically conceived, symbol grounding should be understood in a broadened context in order to fully address the intentionality question, which is at the heart of the matter.

Let us look into some details of this position.

3 Everyday Activities and Symbol Grounding

What is the structure of everyday experience of cognitive agents? How is representation acquired in that experience? Where do its meanings lie?

Let us draw some ideas from phenomenological philosophers such as Martin Heidegger (1927 a, b) and Maurice Merleau-Ponty (1962, 1963) (and also Dewey 1958, Gibson 1950, 1979, Rorty 1979, 1991, Bickhard 1993, Bruner 1995). What is particularly interesting is the notion of *being-in-the-world* (Heidegger 1927, King 1964, Dreyfus 1992, Okrent 1996). The idea is that our existence in the world, or our being-in-the-world, is fundamental to us, to our being what we are. *Being-in-the-world* entails that we constantly interact with the world in a natural, immediate, and non-reflective (i.e., reflexive)

[5] We could, for example, use the word "party" to mean anything from "a political organization" to "an informal gathering", or even "highway construction" (if we decide to interpret it arbitrarily).

[6] Admittedly, though, some unfortunately adopt the ostrich strategy with the belief that if one ignores the argument, it will go away.

[7] This position is similar to, but considerably weaker than, Searle's own view: he thinks that biological systems have some special properties that are the basis of their intentionality, which cannot be captured by computational systems.

way, in our everyday activities in the world. It is believed that such "mindless" everyday activities or coping with the world (on top of our biological pre-endowment) is the basis of our high-level thinking and its intentionality. [8]

According to Heidegger, our everyday coping with the world presupposes a background of common, everyday practices (Dreyfus 1972, 1982, 1992). The "background of practices" (Dreyfus 1992) is not represented, in an explicit and elaborate fashion (such as what we see in traditional symbolic representation, e.g., in the rule base of an expert system; cf. Anderson 1983, Klahr et al 1987), which spells out every detail and every minute variation of every possible situation (Carnap 1969), but is assumed in our *comportment* toward the world. In other words, the most important and fundamental part of our mind is *embodied* and *embedded*, not explicitly represented, and thus it is not directly and explicitly accessible to our (critical) reflection.

Another important theoretical notion is behavioral *structure* (or *form*; Merleau-Ponty 1963, Madison 1981). Maurice Merleau-Ponty extended Heidegger's notions, emphasizing the important role of the structural whole in the understanding of agents. It is not just external situations and internal reactions, but the structural connection that links the two that matters (the affordance and the effectivity, in Gibsonian parlance; Turvey 1992). Situations and reactions are linked on the basis of their shared participation in structures that are comprised of agents and the world (for further discussions, see also Hammond et al 1995, Hutchins 1995, Ballard 1991, Zhang and Norman 1994, Agre and Horswill 1997). [9] According to Merleau-Ponty, in such structures there lies the key to the understanding of meanings or signification of agent's behavior.

3.1 Comportment

Heidegger (1927 a, b) proposed that there is a primordial kind of *comportment* in an agent that directly involves the agent with the world, without the mediation of (explicit) representation. It is a pattern of direct interaction between the agent and its environment, the world. As he put it, "comportment has the structure of directing-oneselves-toward, of being-directed-toward." (Heidegger 1927b). This term is meant to capture the direct two-way interaction of an agent and its world (the "dialectics"; Merleau-Ponty 1963).

Let us explore this notion further to better understand the interaction and the mutual dependency between an agent and its world (at the subconceptual level; Smolensky 1988). First of all, comportment is direct and unmediated. Thus it is free from representationalist baggage. In other words, comportment does not necessarily involve, or presuppose, (explicit) representation and all the problems and issues associated with explicit representation (as discussed earlier). To the contrary, (explicit) representation, and relations between mental states and their objects, presuppose the existence of comportment. Direct and unmediated comportment is in fact the condition of possibility of all representation. Comportment, according to Heidegger (1927a), "makes possible every intentional relation to beings" and "precedes every possible mode of activity in general", prior to (explicit) beliefs, prior to (explicit) knowledge, and prior to (explicit) conceptual thinking (Heidegger 1927a). That is,

[8] The term "everyday activities" as used here does not include *all* everyday activities, but those that are reactive and routine-like. Certain activities that we perform everyday can rely heavily on high-level thinking, for example, if we discuss mathematics everyday. We do not include such exceptions in our definition of everyday activities (as in Dreyfus 1992).

[9] Temporal and spatial patterns (i.e., structures) can be formed in behavior to link various external situations to various reactions and vice versa, and the patterns/structures thus formed determine the essential features of an agent.

comportment is primary. The mistake of the traditional approaches lies in the fact that they treat (explicit) knowledge and its correlates as primary instead, and thus they turn the priority upside-down; in so doing, "every act of directing oneself toward something receives [wrongly] the characteristics of knowing" (Heidegger 1927b). This is in essence the problem of classical cognitivism, including the difficulties with nature of representation, intentionality, consciousness, and so on. On the other hand, in the real world, agents "fix their beliefs not only in their heads but in their worlds, as they *attune* themselves differently to different parts of the world as a result of their experience" (Sanders 1996).

Heidegger's philosophy eschewed the traditional internal/external dichotomy (Bechtel 1988, Pollack 1989) and, in its stead, posited a primordial way of interaction ("comportment"), direct and unmediated, as a foundation upon which cognitive processes (including high-level conceptual thinking) can exist. Being-in-the-world thus serves as a focal point of a different way of thinking about cognition. However, this is not at all to deny the existence of representation. To the contrary, high-level conceptual thinking involving (explicit) representation (as studied extensively in cognitive science and AI) does appear, but it is not as prevalent and as basic as readily assumed in classical cognitivism. It occurs only in more unusual circumstances.

It is useful to point out that comportment that we are talking about is not exactly the same as "embodiment" that has been advocated as the key to understanding human cognition. Lakoff and Johnson (see Lakoff and Johnson 1980 and Johnson 1987) have been putting forth a view that cognition is largely determined by the bodily basis of cognitive agents: The bodily schemata get abstracted and mapped onto all other domains and all other cognitive processes. Interpreted based on that position, an object should not be understood and represented in terms of its shape, color, or any other static features, but should be approached in terms of what an agent can do with it (Glenberg 1997). However, although these ideas are on the right track, 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 coffee, tea, water, powder, coins, paper clips, or business cards, we can hold it in our hands, we can put it on top of our heads, we can stand on it, or we can play tricks with it. Its uses are unlimited. How can an agent structure its understanding around so many different possibilities? The key here, I believe, is what an agent *usually*, *routinely*, *reflexively* do with an object in its everyday life (amidst the "contexture of functionality"), i.e., the common and direct dealing with an object, which is what we call comportment (of an agent) toward an object (i.e., being-with-things; Heidegger 1927a). Such comportment (or everyday routine dealing) with objects is the basis of how an agent approaches objects.

It is also important to note that in our view of comportment, there is no elaborate structure or network that explicitly encodes outcomes, enabling conditions, and other related information (as in Bickhard 1998, which presents an alternative view). In some versions of situated cognition or interactivism (such as Bickhard 1998, Cherian and Troxell 1995), certain elaborate networks of pointers or other mechanisms are devised that relate different internal states in terms of their interrelations in action selection. Although they started out in the right direction, such schemes unfortunately fell right back into the representationalist trap and failed to capture the direct and unmediated nature of comportment. (For possible computational processes that are direct and unmediated, see the next section.)

Direct and unmediated comportment has been variously referred to as reactive routines, reactive skills, routine activities, everyday activities, ongoing interaction, everyday coping, and so on, in the work of e.g. Agre (1988), Chapman (1991), Dreyfus (1991), and Lave (1988). They have also been

classified as subconceptual "representation" (in the broadest sense of the term "representation"; Smolensky 1988, Sun 1994). Brooks (1991), Maes and Brooks (1990), Agre (1988), and Chapman (1991) attempted to implement comportment in a variety of concrete ways, albeit without learning or development (more on learning later).

3.2 Conceptual Processes and Representation

Evidently, comportment is intentional only in the sense that it directs an agent to objects in the world as part of a "natural" structure (Merleau-Ponty 1963). Such intentionality of an agent is not qualitatively different from that of a tennis ball or a car (see the discussion of both earlier). This kind of "pre-representational" (i.e., implicit) pattern of interaction with the world serves as the foundation of how an agent relates to its environment in its everyday activities and, more importantly, also serves as the foundation of more complex forms of intentionality. Explicit representation can only be formed on top of primordial comportment; explicit representation is secondary, or derivative, and its intentional content is *derived* from direct comportment with the world.

As argued in Sun (1997) and Sun et al (1999), there is ample psychological evidence pointing to "bottom-up" learning that goes from comportment to conceptual, symbolic representation. Several lines of research demonstrate that agents can learn complex skilled activities without first obtaining (a large amount of) explicit knowledge (e.g., Berry and Broadbent 1988, Stanley et al 1989, Lewicki et al 1992, Willingham et al 1989, Reber 1989, Karmiloff-Smith 1986, Schacter 1987). In research on *implicit learning*, Berry and Broadbent (1988), Willingham et al (1989), and Reber (1989) demonstrate a *dissociation* between explicitly represented knowledge and performance in a variety of tasks, including dynamic control tasks (Berry and Broadbent 1988), artificial grammar learning tasks (Reber 1989), and serial reaction time tasks (Willingham et al 1989). Berry and Broadbent (1988) argue that agents can learn to perform the task without being provided a priori explicit knowledge and without being able to explicitly verbalize the rules they used to perform the task. This indicates that skills are not necessarily accompanied by explicit knowledge. Willingham et al (1989) similarly demonstrate that performance is not *always* preceded by explicit knowledge in human skill learning, and show that they are not necessarily correlated either. There are indications that explicit knowledge may arise from skilled activities in some circumstances (see Stanley et al 1989). Using a dynamic control task, Stanley et al. (1989) found that the development of explicit knowledge paralleled but lagged behind the development of skilled performance.

Similar claims concerning the development of performance prior to the development of explicit knowledge have surfaced in a number of other research areas and provided additional support for the bottom-up process. *Implicit memory* research (e.g., Schacter 1987) demonstrates a dissociation between explicit and implicit memories in that an agent's performance can improve by virtue of implicit "retrieval" from memory and the agent can be unaware of the process. *Instrumental conditioning* is typically non-verbal and involves the formation of action sequences without explicit knowledge. It may be applied to simple organisms as well as humans (Gluck and Bower 1988). In *developmental psychology*, Karmiloff-Smith (1986) proposes the idea of "representational redescription". During development, low-level implicit knowledge is transformed into explicit representation and thereby made more accessible. This process is bottom-up.

Note that the generation of low-level comportment during ontogenesis is determined by, at least, the following two important factors: (1) Structures in the external world as perceived by the agent,

which in turn depends on current structures in the agent and therefore on the (ontogenetic and phylogenetic) history of the agent/world interaction, and (2) innate "biases", or built-in constraints and predispositions, which also depend on the (ontogenetic and phylogenetic) history of the agent/world interaction. In turn, the generation of high-level structures (i.e., conceptual representation, with symbols) is, to a significant extent, determined by low-level structures, plus sociocultural influences (especially through signs/symbols existing and employed in a given culture). [10]

On this view, high-level conceptual, symbolic representation is rooted, or grounded, in low-level behavior (comportment) from which it obtains its meanings and for which it provides support and explanations. The rootedness/groundedness is guaranteed by the way high-level representation is produced: It is, in the main, extracted out of low-level behavioral structures. Even culturally transmitted symbols have to be linked up, within the mind of an individual agent, with low-level processes in order to be effective.

It is worth noting that conceptual, symbolic representation so formed is in general formed in a functionally relevant way, in relations to everyday activities of agents. In other words, in general it must bear certain existential and/or biological significance to agents and be in the service of agents' activities. The world is of such a high (or even an infinitely high) dimensionality, and thus there can be no totally objective way of perceiving/conceiving it (due to the complexity), except in relation to what an agent has to do with it in everyday activities. Learning symbolic representation on the basis of comportment provides agents with a viable way of basing their conceptual representation on their everyday activities in a functionally relevant way. [11]

The existence of explicit representation, or at least its importance, has been denied, or downplayed, by many advocates of (strong forms of) situated cognition and interactivism (e.g., Suchman 1987, Brooks 1991, Bickhard 1993, Port and van Gelder 1995). The existence of explicit representation (but not its paramount role) has in fact been argued for by a number of researchers, persuasively, I believe. See e.g. Markman and Dietrich (1998) and Smith et al (1992). Here I take the more "eclectic" position that acknowledges representation is important while maintaining that it is mediated by direct comportment.

3.3 A Dual Process Theory

This analysis boils down to the dual process theory (i.e., the *dual-level hypothesis*; Sun 1994). In one of my previous books (entitled *Integrating Rules and Connectionism for Robust Commonsense Reasoning*; Sun 1994), I put forth the following hypothesis (in which the word "knowledge" should be broadly interpreted):

> It is assumed in this work that cognitive processes are carried out in two distinct levels with qualitatively different processing 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.

[10] Culture also has the role of structuring (constraining) the interaction of an individual agent with the world through the mediating tools, signs, and other cultural artifacts, and thus it affects low-level structures too, although to a lesser extent. In contrast to Vygotsky (1962), though, I would emphasize equally internally generated signs/symbols and externally transmitted ones.

[11] In addition, of course, biological pre-endowment in agents (acquired through evolutionary processes) may also provide them with some ways of picking out relevant information. The two aspects may interact closely in forming conceptual representation.

This idea is closely related to some well-known dichotomies in cognitive science: the dichotomy of symbolic vs. subsymbolic processing (Rumelhart and McClelland 1986), the dichotomy of conceptual vs. subconceptual processing (Smolensky 1988), the dichotomy of explicit vs. implicit learning (Reber 1989, Berry and Broadbent 1988, Lewicki et al 1992), the dichotomy of controlled vs. automatic processing (Shiffrin and Schneider 1977, Schneider and Oliver 1991), and the dichotomy of declarative and procedural knowledge (Anderson 1983). However, different from some of these dichotomies, I went further in positing separate and simultaneous existence of multiple levels (i.e., separate processors), each of which embodies one side of a dichotomy (cf. Grossberg 1987). Therefore, in this work, the two sides of a dichotomy are not simply two ends of a spectrum, or two levels of analysis of the same underlying system. But they are two separate, although closely connected, systems.

In the current context, the two distinct levels are termed, respectively, *the conceptual level* and *the subconceptual level*, following the usage by Smolensky (1988). The two levels encode similar and comparable content (or "knowledge" in a broad sense). But they encode their content in different ways. One works in a comportment-like way while the other in an explicit, symbolic, and conceptual way. They also utilize different processing mechanisms. Thus, they can have qualitatively different flavors, although they can function together. The reason for having the two levels, or any other similar combination of components, is that these different levels can potentially work together *synergistically*, supplementing and complementing each other in a variety of different ways (as demonstrated in Sun 1994, 1997, Sun and Peterson 1998 a, b). [12] I have argued for the two-level hypothesis extensively before, based on a variety of evidence (see, e.g., Sun 1995, 1997). So I will not repeat the arguments here.

4 Computational Analysis of Everyday Activities

Let me briefly sketch a picture of the mechanistic underpinning of this theory, that is, an architecture for a model of the mind. [13] I want to put together here some basic ingredients. First of all, the new approach should start small, with only minimum built-in initial structures in an agent. Some of these initial structures have to do with "pre-wired" reflexes, or predisposition for developing such reflexes, that is, genetic and biological pre-endowment. Some others have to do with learning capabilities, since most of the structures in cognition will have to be constructed in a gradual, incremental fashion, during the course of individual ontogenesis (so one might view this as a constructivist approach, although different from the Piagetian approach; cf. Inhelder and Piaget 1958). The development of structures is based on interaction with the world (i.e., based on being-in-the-world, including both the physical world and the social/cultural world). The interaction prompts the formation of various low-level structures in behavior, which in turn lead to the emergence of high-level conceptual representation.

In developing this approach, connectionist models are utilized, in a properly generalized form, as the basic unifying medium in implementation, and as a guiding metaphor for constructing hypotheses, models, and theories (Sun 1994, 1995, Sun and Bookman 1994), because of the many appealing properties of such models (Smolensky 1988, Bechtel and Abrahamsen 1991, Clark 1993) and because of the fact that it encompasses both symbolic approaches and dynamical system approaches (Port and van Gelder 1995).

[12] See Sun (1997) and Sun et al (1999) regarding this synergy hypothesis.

[13] This is what I believe to be needed for an integrative cognitive science that seeks (1) to instantiate its theories and (2) to fit various pieces together to form a coherent whole.

4.1 Computational Processes of Comportment

Now we are ready to probe these ideas further. First, we need to gain a better understanding of comportment, beyond mere philosophical speculation. In this subsection, we shall examine the computational processes of comportment. We shall also look into the *development* of comportment in the "ontogenesis" of an individual agent, which is the most important means by which comportment is acquired (although some innate structures might be formed evolutionarily, a priori, as mentioned before).

Generally speaking, while performing everyday activities (especially in direct comportment), the agent is under time pressure: Often, a mundane action "decision" has to be made in a fraction of a second; it cannot involve much of "information processing", and falls outside of Allen Newell's "rational band" (i.e, cognitive processes that take minutes or hours to complete, which is what cognitive science and AI traditionally deal with; Rosenbloom et al 1993). The agent is also severely limited in other resources, such as memory, so that memorizing and analyzing all the previous experience (in detail) is impossible (although some form of episodic memory obviously exists). The perceptual ability of the agent is also limited in that only local information is available. Goals may not be explicit and a priori to an agent either. They may be implied in reinforcements/payoffs received, and they are pursued by an agent as a side-effect of trying to maximize reinforcements/payoffs.

Learning of comportment is an experiential, trial-and-error process; the agent develops its competence, *tentatively*, on an on-going basis (because it cannot wait until the end of its experience before making a decision and starting to learn). [14] In general, as demonstrated by the models of Nosofsky et al (1994) and Medin et al. (1987), human learning is mostly gradual, on-going, and concurrent (on-line), which is especially true in learning comportment. The characteristics of the world, from the viewpoint of the learning agent, need not be stationary. It can be nonstationary ("drifting") in several ways: (1) The world can change over time; thus, the revision of comportment structures learned by an agent may be necessary. (2) Even when the world per se is stationary, it may still seem evolving to an agent learning to cope with the world, because different regions of the world may exhibit different characteristics and thus revisions over time may be required (Widmer and Kubat 1996). In general, there is no preselected set of instances that provide a fixed view of the world. (3) Once a structure is revised, the agent has to view whatever it experienced before in new ways (because the current comportment structures serve as a "filter" through which the agent sees the world), and thus the experience may seem different and the world nonstationary. (4) There is a lack of a clear and steady criterion for learning comportment. Reinforcement/payoffs may be received sporadically, and it is up to the agent to decide what to make of them. The agent has to assign credits/blames on the basis of what is already known, which is constantly changing. Since the learning criterion is a moving target, the learning process becomes nonstationary.

There are some existing computational methods available to accomplish simple forms of such learning. Chief among them is the temporal difference method (Sutton 1988), a type of reinforcement learning that learns through exploiting the difference in evaluating actions in successive steps and thus handling sequences in an incremental fashion. Another approach, genetic algorithm (Holland et al 1986), may also be used to tackle this kind of task. An an example, to implement the first approach using neural nets, as discussed in Sun (1997), we can use a four-layered network in which the first three layers form a backpropagation network (feedforward or recurrent; Rumelhart and McClelland

[14] It may not be clear what constitutes a unit of experience and how long it is; resource limitation may prevent the agent from remembering sequences of past events.

1986) for computing output action values (i.e., Q-values; Watkins 1989) and the fourth layer performs stochastic decision making (Watkins 1989). The network can be internally subsymbolic, involving distributed features as developed automatically through the backpropagation algorithm. The output of the third layer indicates the value of each action (represented by an individual node). The value is an evaluation of the "quality" of an action in a given input state. To acquire these values, we can use the standard *Q-learning* algorithm (a temporal difference reinforcement learning algorithm as mentioned earlier). It basically compares the values of successive actions and adjusts an evaluation function on that basis, without explicitly involving and representing goals, states, and outcomes. (For details of Q-learning, see Watkins (1989) as well as more recent treatments such as Kaelbling et al (1996).) The afore-specified model is in fact one part of our overall model (named CLARION), that is specifically concerned with comportment (Sun 1997, Sun and Peterson 1998 a, b). Clearly, in this implementation of comportment, there is no elaborate internal representation, and no pointer between different entities representing mental states (cf. Bickhard 1998), as has been specified earlier.

4.2 Computational Processes of Conceptual Processing

Let us now discuss computational processes of high-level conceptual processing with (explicit) representation, during everyday activities of agents.

Let us see how explicit representation is acquired. We have discussed the idea of "bottom-up" learning. But how do we accomplish "bottom-up" learning computationally? Admittedly, there are many symbolic rule learning algorithms out there for learning explicit rules (as a form of high-level explicit representation). However, the afore-identified characteristics of everyday activities (section 4.1) render most existing rule learning algorithms inapplicable, because they require either preconstructed exemplar sets (Michalski 1983, Quinlan 1986, 1990), incrementally given consistent instances (Mitchell 1982, Fisher 1986, Utgoff 1989), or complex manipulations of learned symbolic structures when inconsistency is discovered (which is typically more complex than the limited time an agent may have in reactive activities; Hirsh 1994). "Drifting" as analyzed before is clearly more than noise and inconsistency as considered by some learning algorithms, because it involves changes over time and may lead to radical changes in learned knowledge. Above all, most of the rule learning algorithms do not handle the learning of sequences, which is an essential form of agents' everyday activities and necessarily involves temporal credit assignment.

Therefore, algorithms have to be developed specifically for the purpose of modeling the acquisition of explicit representation in agents (Sun and Peterson 1998 a, b). The algorithms should be bottom-up; that is, they utilize whatever is learned in the lower level (the part of the computational model that implements comportment, described in the previous subsection) and construct symbolic representation at the higher level (the part that implements conceptual processing in agents). The basic idea for such bottom-up learning is as follows: If some action decided by the bottom level of an agent is successful (being successful could mean a number of different things; see Sun and Peterson 1998 a, b regarding details), then the agent extracts an explicit rule that corresponds to the action selected 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 and exclusive of the current case; if the outcome is successful, the agent may try to generalize the rule to make it more universal. (Specialization and generalization are done based on actually encountered situations, and thus search is minimal.) To measure the success or failure of a step in order to learn rules, some statistical criteria (based on

information gain) have been developed and tested in Sun and Peterson (1998a, b).

We use a localist connectionist model for representing these rules (where "localist" means that each concept is represented separately by an individual node in a network). Basically, we connect the nodes representing conditions of a rule to the node representing the conclusion (through standard sigmoid activation functions). That is, we translate the structure of a set of rules into the structure of a network. See e.g. Sun (1994, 1995) for details of localist encoding. The rule learning algorithm (as described above) is implemented on top of the localist network (Sun and Peterson 1998 a, b). The use of the learned rules is through a number of inference algorithms (such as backward chaining and forward chaining), which are also implemented on top of the network (through controlling activation propagation; Sun 1997).

Besides bottom-up influences, there are also "top-down" influences (in learning and otherwise). In learning, such influences have been amply studied and modeled by top-down learning models such as ACT (Anderson 1983 and Anderson and Lebiere 1998). In our framework, when a rule is given, the agent can represent it in the top-level rule network and connect it to existing representation. [15] We can go one step further by performing rule assimilation, by which rules given externally, besides being represented in the top level, are assimilated into the bottom level and thus become more effective (see Anderson 1983, Dreyfus and Dreyfus 1987). We can train the bottom level according to rules given in the top level, using supervised learning (e.g., backpropagation). Aside from learning, top-down influences are apparent in performance as well. For example, Rips (1989) showed that even categorization does not rely on just similarity but also reasoning from rules. The computational processes of top-down influences can be based on combining the outcomes of the two levels in making decisions (Sun 1997).

The afore-described ideas concerning computational processes of both comportment (subconceptual processing) and conceptual processing have been implemented in a (computational) cognitive model named CLARION. The model has been described extensively in a series of papers, including Sun (1997), Sun and Peterson (1998 a, b), and Sun et al (1998a, b, 1999). Essentially, as described above, it is made up of two levels, the top level conceptual and the bottom level subconceptual (comportment-oriented). [16] The two levels interact in action selection, through combining action recommendations from the two levels respectively, and they cooperate in learning through the afore-described bottom-up and top-down processes. The model grounds symbols and symbolic representation through bottom-up learning and through comportment with the world.

4.3 Concept Formation

While learning rules, CLARION forms concepts (Sun and Peterson 1998 b). Although there are available distributed features in the bottom level for specifying rule conditions, a separate node is instead set up in the top level to represent the condition of a rule, that connects to the distributed features. So along with induced rules, localist (i.e., symbolic) representation of concepts are formed. Each localist node is linked to its corresponding distributed features in the bottom level, from which it was abstracted, through bottom-up activation (using a standard sigmoid activation function with uniform weights).

This kind of concept representation is basically a *prototype model* (Smith and Medin 1981, Rosch

[15] Alternatively, supervised learning on the rule network can be performed with, e.g., backpropagation for slower learning of the rule (Rumelhart and McClelland 1986).

[16] There is also a separate memory system as described in Sun (1997).

1978). Localist nodes serve as identification and encoding of features, in a bottom-up direction. They also serve to trigger relevant distributed features, in a top-down direction, once a concept is brought into attention. [17]

Moreover, concepts formed in this way in CLARION are context-dependent and task-oriented, because they are always formed with regard to the tasks at hand while exploiting environmental regularities (Heidegger 1927a, Okrent 1996). The representation of an agent, for the most part, need not be determined a priori. Being emphasized in CLARION is the functional role of concepts and the importance of function in forming concepts. A concept is formed as part of a rule, which is learned to accomplish a task in a certain environment. Therefore, acquired concepts are functional. The task context and the experience help an agent to determine which features in the environment need to be emphasized, what objects should be grouped together, and thus what constitutes a separate category of objects (that is, a concept). (See Sun and Peterson (1998 b) for a more detailed analysis of concepts formed by agents during learning of specific tasks.) This may explain why most human concepts are (more or less) concerned with existentially and biologically significant aspects of the world; they are not just "objective" classifications (Lakoff and Johnson 1980, Lave 1988).

Such concept formation and representation may have interesting implications for the frame problem in AI. The frame problem refers to the difficulty in keeping track of many propositions concerning a situation that may be in a constant flux. Any change in the situation may lead to the validation of many propositions and the invalidation of many others. This may lead to very high computational complexity when we need to reason about the situation. This idea of "tracking" was envisaged in the purely symbolic representational frameworks of traditional AI, which make such a process necessary. However, concepts being grounded, context-dependent, and task-oriented (such as in CLARION) may alleviate the need for such "tracking". In this alternative approach, each situation is reacted to, first and foremost, by low-level comportment and then, through low-level comportment, it triggers proper concepts and propositions at the higher level, which thus produces inferences that are tailored to the situation (Sun 1994). Purely logical reasoning concerning each and every proposition possible (as envisaged by the traditional approaches) is computationally excessive and is rendered unnecessary by this approach (Sun 1994, 1995). (Certainly, we need much more work along this direction to further elucidate the potential of this alternative view.)

5 Representation and Intentionality: An Assessment

Now we can revisit the issue of representation. I was critical of the traditionally dominant and currently lingering position on representation in cognitive science. However, rejecting representationalism does not necessarily mean rejecting (weak notions of) representation and symbols.

Let us examine CLARION. In this model, the top level indeed consists of representation in the sense implied by representationalism; this is because encoding used there is punctate (with each item being in an isolatable node) and also elaborate (with represented items forming a rather complete model of a domain, if CLARION is given enough time to learn). Moreover, the representation is symbolic, in the sense that a concept is assigned to an arbitrary node, without any intrinsic connection between a node and what it represents. Syntactic structures (concatenative compositional structures; Fodor 1975) can be built on the basis of such representation. Syntactically sensitive symbolic processing

[17] They also facilitate "inheritance" reasoning using distributed features, as discussed in (Sun 1994, 1995).

can be performed on them. This level of the model is thus representational. However, the bottom level of CLARION is different. A connectionist model (viz. a backpropagation network; Rumelhart and McClelland 1986) is used; thus distributed feature encoding is involved. In such a scheme, there is no symbol, that is, no lexical item that can be assigned arbitrary meanings. Moreover, it is not a priori determined (beside certain presumably biologically built-in constraints). The encoding does not exist before an agent learns to interact with the world and thus develops; it is intrinsically tied to the experience of interaction between an agent and its environment. There is no syntactic structure in that level (in the sense of concatenative compositional structures; Fodor 1975). Thus this level of CLARION is non-representational. Putting the two levels together, CLARION incorporates both representation and nonrepresentation. However, the model does not simply juxtapose the two qualitatively different components, but combines them in an integral framework, so the two parts of the model can interact and cooperate. [18]

What is the implication of the above discussion for the question of intentionality? Let us compare how the meanings of the contents of the bottom level and the top level are determined. To understand this issue, the notions of intrinsic intentionality and derived intentionality (Searle 1980) are pertinent. According to Heidegger (1927b), representation presupposes the more basic comportment with the world. Comportment carries with it a direct and unmediated relation to, and reference of, things in the world; being-with-things is a fundamental form of being-in-the-world, a bridge to the existential context of an agent. Therefore, it provides an *intrinsic* intentionality (meanings), or in other words, a connection (to things in the world) that is intrinsic to an agent, given a particular existential context of the agent (and its biological pre-endowment). In addition to intrinsic intentionality, there is the other kind of intentionality, derived intentionality, which is obtained through derivative means. In CLARION, intentionality can be categorized into these two kinds precisely: the bottom-level processes that capture direct comportment with the world and the top-level processes that are the result of extracting rules and concepts from the bottom-level processes. The bottom-level processes acquire their internal encoding through learning from the experience of direct interaction with the world, and thus the meanings of the encoding lie in the intrinsicness of the weights and the wiring, which are determined by the process of unmediated interaction (using Q-learning and backpropagation in the model). [19] The top-level processes result from extraction, i.e., derivation. Thus the meanings or intentionality of the representation can only be traced to the derivation process. Through derivation/extraction, as well as through the on-going connection to the bottom level (both top-down and bottom-up connections), symbols at the top level are "grounded" in the bottom level and, through the bottom level, in the comportment with the world. I want to emphasize that not only symbols themselves are derived from the bottom-level processes but the meanings of these symbols are thus also derived from these processes.

Moreover, although Heidegger recognized the ontological precedence of intrinsic intentionality, it is also important to further recognize, as in our model, that intrinsic intentionality is not only ontologically prior to derived intentionality, but also developmentally (ontogenetically) prior to it, in individual agents. As demonstrated by Inhelder and Piaget (1958), a child learns concepts and schemas only when the child learns to interact with objects in the world, and more and more complex concepts are developed through more and more complex interaction with objects. As suggested by Karmiloff-Smith (1992), the increasing sophistication and mastery of concepts are accomplished, in part, through

[18] This cooperation produces synergistic results (see Sun and Peterson 1998 a, b for demonstrations of synergy resulting from the interaction of the two components).

[19] Such weights and wiring, unlike arbitrarily selected encoding at the top level, are intrinsically determined by the input and output during the interaction, as well as by their initial settings.

a "representational redescription" process: First, a child acquires an embodied performance ability, then through representational redescription, i.e., extracting explicit representation, the child learns explicit concepts and thereby further improves performance. CLARION, as described earlier, roughly captures this developmental process (in a qualitative way; Sun 1997).

6 Further Discussions

6.1 Comparisons

We can contrast the above-outlined approach with some traditional thinking on cognition. One major difference we see is that traditional thinking tends to overlook various external factors in cognition. David Hume (Hume 1938) believed that cognition can only be understood on the basis of sense data that an individual agent receives, based on which an agent forms associations that relate them, in order to make sense of them. In William James's *The Principles of Psychology* (James 1890), in a total of 28 chapters covering a wide ranging set of topics, cognition is construed as merely an internal process that works on data provided by external sources. In *Readings in Cognitive Science* edited by Collins and Smith (1988), a major collection of significant early work in cognitive science, the field of cognitive science was defined to be "the interdisciplinary study of the acquisition and use of knowledge". Despite the fact that knowledge is the result of agents' interaction with the world (individually and/or collectively), there is no treatment of such dynamic on-going interaction in the book.

In contemporary cognitive science and AI, although ideas similar to some of those outlined in the present framework have started to seep into various segments of different research communities (see, e.g., Russell and Norvig 1995, Hutchins 1995, Agre 1995, Damasio 1994, Sun 1994), cognitive science/AI, as a whole, has not been particularly hospitable to these new ideas. See e.g. Vera and Simon (1993) and Hayes, Ford and Agnew (1994) for a glimpse of the opposing views.

Another contrast is with regard to the duality in cognition (the dual processes), which has long been speculated upon. Although the notion of the conscious vs. the subconscious has captivated pop culture ever since Freud (Freud 1937, Kitcher 1992), in mainstream academic psychology (especially cognitive psychology) and in mainstream academic philosophy, it is not quite readily accepted (although there are some notable exceptions). The distinction of the conceptual vs. the subconceptual was proposed, in the context of analyzing connectionist models (Smolensky 1988), as a sort of substitute for the distinction of the conscious vs. the subconscious (to avoid the controversies surrounding the latter). The distinction of the conceptual vs. the subconceptual has not been very popular either. In contrast, we take such dichotomies seriously and use them as the basis of our approach.

One additional note should be made regarding the relation between our dichotomy and the dichotomy of declarative vs. procedural knowledge (Anderson 1983), which came closest to our dichotomy. The two dichotomies are very similar, except that Anderson's model did not account well for bottom-up learning, because it is based (mostly) on top-down learning (to capture various instructed learning situations), and thus it did not account for the derivation and the grounding of symbols and representation.

The approach outlined here is consistent with the situated cognition view (interactivism), in the sense that coping with the world means acting in an environmentally driven fashion and dealing with

moment-to-moment contingencies. Our approach reflects such a view through a focus on reacting to the current state of the world. Also in line with the situated cognition view and interactivism, learning in our approach is tied closely to specific situations as experienced (to reflect and exploit environmental contingencies and regularities). But there are some obvious differences. The situated cognition view often claims that there should not be any elaborate model of the world or elaborate representation. However, instead of being completely antithetical to the representationalist view and hastily avoiding any representation or model, we take a more inclusive approach: We show that explicit representation *can* be constructed on the basis of situated learning by situated agents through a bottom-up process, thus unifying the two contradictory views.

6.2 Concluding Remarks

This article shows how representation and representational content emerge in the interaction between an agent and its environment. It hypothesizes the process that goes from unmediated comportment with the world to mediated (symbolic and conceptual) representation and the concomitant conceptual processing.

Our framework outlined above reconciles representationalism and situated cognition interactivism. It does so through explicating the crucial role played by direct and unmediated comportment. Comportment bridges the gap between the world and the internal (explicit) representation in an agent. It makes (explicit) representation possible by giving rise to it through experience.

The key of our work lies in recognizing the fact that experiences (everyday activities) come first and representation comes later (which has been recognized by many) and, based on that , constructing computational models that demonstrate the feasibility of this view (which is novel). Instead of Descartes' motto "I think, therefore I am" (or the revisionist version "I feel, therefore I am"; Damasio 1994), we now argue that (computationally) it can be and should be "I am, therefore I think". I believe that this reversal sets the priority right for computational cognitive modeling, as well as for cognitive science in general.

References

[] P. Agre, (1988). The Dynamic Structure of Everyday Life, Technical Report, MIT AI Lab, Cambridge, MA.

[] P. Agre, (1995). Computational research on interaction and agency. *Artificial Intelligence*. Vol.72, 1-52.

[] P. Agre and I. Horswill, (1997). Lifeworld analysis. *Journal of Artificial Intelligence Research*. Vol.6, 111-145.

[] J. R. Anderson, (1983). *The Architecture of Cognition*, Harvard University Press, Cambridge, MA

[] J. R. Anderson, (1985). *Cognitive Psychology and Its Implications*. W.H.Freeman, New York, NY.

[] J. Anderson and C. Lebiere, (1998). *The Atomic Components of Thought*, Lawrence Erlbaum Associates, Mahwah, NJ.

[] D. Ballard, (1991). Animate vision. *Artificial Intelligence*. Vol.48, pp.57-86.

[] L. Barsalou, (1989). Intraconcept similarity and its implications. in S. Vosniadou and A. Ortony (eds.), (1989) *Similarity and Analogical Reasoning*, Cambridge University Press, New York.

[] L. Barsalou, (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22.

[] W. Bechtel, (1988). *Philosophy of Mind: an Overview for Cognitive Science.* Lawrence Erlbaum and Associates, Hillsdale, NJ.

[] W. Bechtel and A. Abrahamsen, (1991). *Connectionism and the Mind.* Basil Blackwell, Oxford, UK.

[] W. Bechtel and G. Graham, (eds.) (1998). *A Companion to Cognitive Science.* Blackwell, Oxford, UK.

[] D. Berry and D. Broadbent, (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology.* 79, 251-272.

[] M. Bickhard, (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence.* pp.285-333.

[] M. Bickhard, (1998). Interaction and Representation. http://www.lehigh.edu/ bickhard

[] R. Brooks, (1991). Intelligence without representation. *Artificial Intelligence.* Vol.47. pp.139-159.

[] R. Brooks and L. Stein, (1994). Building brains for bodies. *Autonomous Robots.* Vol.1, No.1. pp.7-26.

[] J. Bruner, (1995). *Acts of Meaning.* Harvard University Press, Cambridge, MA.

[] R. Carnap, (1969). *The Logic Structure of the World.* University of California Press.

[] D. Chalmers, (1989). Why Fodor and Pylyshyn were wrong: the simplest refutation. *Proceedings of Cognitive Science Conference.* Lawrence Erlbaum Associates, Hillsdale, NJ.

[] D. Chapman, (1991). *Vision, Instruction, and Action.* MIT Press, Cambridge, MA.

[] S. Cherian and W. Troxell, (1995). Intelligent behavior in machines emerging from a collection of interactive control structures. *Computational Intelligence*, 11 (4), 565-592.

[] P. Churchland, (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* MIT Press, Cambridge, MA.

[] A. Clark, (1993). *Associative Engines: Connectionism, Concepts, and Representatinal Change.* MIT Press, Cambridge, MA.

[] A. Collins and E. Smith, (1988). *Readings in Cognitive Science.* MIT Press, Cambridge, MA.

[] A. Damasio, (1994). *Decartes' Error: Emotion, Reason, and the Human Brain.* Grosset/Putnam, New York.

[] E. Davis, (1990). *Representations of Commonsense Knowledge.* Morgan Kaufman, San Mateo, CA.

[] J. Dewey, (1958). *Experience and Nature.* Dover, New York.

[] F. Dretske, (1981). *Knowledge and the Flow of Information.* MIT Press, Cambridge, MA.

[] H. Dreyfus, (1972). *What Computers Can't Do.* Harper and Row. New York.

[] H. Dreyfus, (1982). *Husserl, Intentionality, and Cognitive Science*, MIT Press, Cambridge, MA.

[] H. Dreyfus and S. Dreyfus, (1987). *Mind Over Machine: The Power of Human Intuition.* The Free Press, New York.

[] H. Dreyfus, (1992). *Being-in-the-world.* MIT Press, Cambridge, MA.

[] H. Dreyfus, (1992). Introduction to the MIT Press Edition, In: *What Computers Still Can't Do*, MIT Press, Cambridge, MA.

[] D. Fisher, (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*. 2, 139-172.

[] J. Fodor, (1975). *The Language of Thought*. Crowell.

[] J. Fodor and Z. Pylyshyn, (1988). Connectionism and Cognitive Architecture: A Critical Analysis, in: Pinker and Mehler (eds.) *Connections and Symbols*, MIT Press, Cambridge, MA. 1988

[] W. Freeman, (1995). *Societies of Brains: A Study in the Neuroscience of Love and Hate*. Lawrence Erlbaum Associates, Hillsdale, NJ.

[] S. Freud, (1937). *A General Selection from the Works of Sigmund Freud*. Hogarth Press. London, UK.

[] T. van Gelder, (1990). Compositionality: a connectionist variation on a classical theme. *Cognitive Science*. Vol.14, 355-384.

[] J. Gibson, (1950). *The Perception of the Visual World*. Houghton Mifflin. Boston, MA.

[] J. Gibson, (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin. Boston, MA.

[] A. Glenberg, (1997). What memory is for. *Brain and Behavioral Sciences*.

[] M. Gluck and G. Bower, (1988). From conditioning to category learning. *Journal of Experimental Psychology: General*. 117 (3), 227-247.

[] S. Grossberg, (1987). *The Adaptive Brain*. North-Holland, New York, NY.

[] K. Hammond, T. Converse, and J. Grass, (1995). The stabilization of environments. *Artificial Intelligence*. Vol.72, 305-328.

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

[] P. Hayes, K. Ford, and N. Agnew, (1994). On babies and bathwater: a cautionary tale. *AI Magazine*. Vol.15, No.4, pp.14-26.

[] M. Heidegger, (1927a). *Being and Time*. English translation published by Harper and Row, New York. 1962.

[] M. Heidegger, (1927b). *The Basic Problem of Phenomenology*.

[] H. Hirsh, (1994). Generalizing version spaces. *Machine Learning*, 17, 5-46.

[] J. Holland, N. Nisbitt, T. Thagard and J. Holyoak, (1986). *Induction: A Theory of Learning and Development*. MIT Press, Cambridge, MA.

[] D. Hume, (1938). *An Abstract of A Treatise of Human Nature*. Cambridge University Press

[] E. Husserl, (1970). *Logical Investigation*. London, Routledge and K. Paul; New York, Humanities Press

[] E. Hutchins, (1995). How a cockpit remembers its speeds. *Cognitive Science*. 19, 265-288.

[] B. Inhelder & J. Piaget, (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. Routledge & Kegan Paul, London, England.

[] W. James, (1890). *The Principles of Psychology*. Dover, New York.

[] M. Johnson, (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Univeristy of Chicago Press, Chicago.

[] L. Kaelbling, M. Littman, and A. Moore, (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237-285.

[] A. Karmiloff-Smith, (1986). From meta-processes to conscious access: evidence from children's metalinguistic and repair data. *Cognition*. 23. 95-147.

[] A. Karmiloff-Smith, (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press, Cambridge, MA.

[] F. Keil, (1989). *Concepts, Kinds, and Cognitive Development*. MIT Press. Cambridge, MA.

[] M. King, (1964). *Heidegger's Philosophy*. Macmillan, New York.

[] P. Kitcher, (1992). *Freud's Dream*. MIT Press, Cambridge, MA.

[] D. Klahr, P. Langley, and R. Neches, (eds.) (1987). *Production System Models of Learning and Development*. MIT Press, Cambridge, MA.

[] G. Lakoff and M. Johnson, (1980). The metaphoric structure of human conceptual system. *Cognitive Science*, Vol.4, pp.193-208.

[] J. Lave, (1988). *Cognition in Practice*. Cambridge University Press, New York.

[] G. Madison, (1981). *The Phenomenology of Merleau-Ponty*. Ohio University Press, Athens, Ohio.

[] P. Maes and R. Brooks, (1990). Learning to coordinate behaviors. *Proc.of AAAI-90*. pp.796-802. Morgan Kaufmann, San Mateo, CA.

[] A. Markman and E. Dietrich, (1998). In defense of representation as mediation. *Psycoloquy*, 9 (48).

[] D. Medin, W. Wattenmaker, and R. Michalski, (1987). Constraints and preferences in inductive learning: an experimental study of human and machine performance. *Cognitive Science*. 11, 299-339.

[] M. Merleau-Ponty, (1962). *Phenomenology of Perception*. Routledge and Kegan Paul, London.

[] M. Merleau-Ponty, (1963). *The Structure of Behavior*. Beacon Press, Boston.

[] R. Michalski, (1983). A theory and methodology of inductive learning. *Artificial Intelligence*. Vol.20, pp.111-161.

[] T. Mitchell, (1982). Generalization as search. *Artificial Intelligence*, 18, 203-226.

[] M. Minsky, (1983). A framework for representing knowledge. In: *Mind Design*. MIT Press. Cambridge, MA.

[] A. Newell and H. Simon, (1976). Computer science as empirical inquiry: symbols and search. *Communication of ACM*. 19. 113-126.

[] M. Okrent, (1996). Why the mind isn't aprogram (but some digital computer might have a mind). *Electronic Journal of Analytic Philosophy*, 4, 1-30. http://www.phil.indiana.edu/ejap/

[] D. Osherson and H. Lasnik, (1990). *An Invitation to Cognitive Science*. MIT Press, Cambridge, MA.

[] C. Peirce, (1955). *The Philosophical Writings of Charles Peirce*. ed. J. Buchler. Dover, New York.

[] S. Pinker, (1994). *The Language Instinct*. W. Morrow and Co. New York, NY.

[] J. Pollack, (1989). *How to Build a Person*. MIT Press, Cambridge, MA.

[] R. Port and T. van Gelder, (1995). *Mind as Motion: Dynamics, Behavior, and Cognition*. MIT Press. Cambridge, MA.

[] M. Posner, (ed.) (1989). *Foundations of Cognitive Science*. MIT Press, Cambridge, MA.

[] H. Putnum. 1975. The meaning of 'meaning'. In K. Gunderson (ed.) *Mind and Knowledge.* University of Minnesota Press. Mpls, MN.

[] R. Quinlan, (1986). Inductive learning of decision trees. *Machine Learning.* 1, 81-106.

[] R. Quinlan, (1990). Learning logical definition from relations. *Machine Learning.* 5, 239-266.

[] A. Reber, (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General.* 118 (3), 219-235.

[] R. Rorty, (1979). *Philosophy and the Mirror of Nature.* Princeton University Press. Princeton, New Jersey.

[] R. Rorty, (1991). *Essays on Heidegger and Others.* Cambridge University Press. New York.

[] E. Rosch, (1978). Principles of categorization. In: E. Rosch and B. Lloyd, (eds.) *Concepts and Categorization.* Lawrence Erlbaum and Associates, Hillsdale, NJ.

[] D. Rumelhart and J. McClelland, eds. (1986). *Parallel Distributed Processing I.* MIT Press, Cambridge, MA.

[] S. Russell and P. Norvig, (1995). *Artificial Intelligence: A Modern Approach.* Prentice Hall, Englewood Cliffs, NJ.

[] J. Sanders, (1996). An ecological approach to cognitive science. *Electronic Journal of Analytic Philosophy.* 4, 1-12. http://www.phil.indiana.edu/ejap

[] W. Schneider and W. Oliver (1991), An instructable connectionist/control architecture. In: K. VanLehn (ed.), *Architectures for Intelligence*, Erlbaum, Hillsdale, NJ.

[] Searle, J (1983). *Intentionality.* Cambridge University Press, New York.

[] Searle, J (1980). Minds, brains, and programs. *Brain and Behavioral Sciences.* 3. 417-457.

[] D. Schacter, (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 13, 501-518.

[] T. Shallice, (1972). Dual functions of consciousness. *Psychological Review.* Vol.79, No.5, 383-393.

[] R. Shiffrin and W. Schneider, (1977). Controlled and automatic human information processing II. *Psychological Review.* 84. 127-190.

[] E. Smith, C. Langston, and R. Nisbett, (1992). The case for rules in reasoning. *Cognitive Science*, 16, 1-40.

[] E. Smith & D. Medin, (1981). *Categories and Concepts.* Cambridge, MA: Harvard University Press.

[] P. Smolensky. (1988). On the proper treatment of connectionism, *Behavioral and Brain Sciences*, 11, 1-43.

[] W. Stanley, R. Mathews, R. Buss, and S. Kotler-Cope, (1989). Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology.* 41A (3), 553-577.

[] L. Suchman, (1987). *Plans and Situated Actions.* Cambridge Univeristy Press, Cambridge, UK.

[] R. Sutton, (1988). Learning to predict by the methods of temporal difference. *Machine Learning*, 3, 9-44.

[] R. Sun, (1994). *Integrating Rules and Connectionism for Robust Commonsense Reasoning.* Wiley, New York.

[] R. Sun, (1995). Robust Reasoning: Integrating Rule-Based and Similarity-Based Reasoning. *Artificial Intelligence*, 75, 2. 241-296.

[] R. Sun, (1997). Learning, action, and consciousness: a hybrid approach towards modeling consciousness. *Neural Networks*, special issue on consciousness. 10 (7), pp.1317-1331.

[] R. Sun and L. Bookman, (eds.) (1994). *Computational Architectures Integrating Neural and Symbolic Processes*. Kluwer Academic Publishers. Norwell, MA.

[] R. Sun and T. Peterson, (1998 a). Some experiments with a hybrid model for learning sequential decision making. *Information Sciences*. Vol.111, 83-107.

[] R. Sun and T. Peterson, (1998 b). Autonomous learning of sequential tasks: experiments and analyses. *IEEE Transaction on Neural Networks*. Vol.9, No.6, pp.1217-1234.

[] R. Sun, E. Merrill, and T. Peterson, (1998 a). A bottom-up model of skill learning. *Proc.of 20th Cognitive Science Society Conference*, pp.1037-1042, Lawrence Erlbaum Associates, Mahwah, NJ.

[] R. Sun, E. Merrill, and T. Peterson, (1998 b). Skill learning using a bottom-up hybrid model. *Proc. of The Second European Conference on Cognitive Modeling*, Nottingham, UK. April, 1998. pp.23-29. Nottingham University Press. Nottingham, UK.

[] R. Sun, E. Merrill, and T. Peterson, (1999). A model of bottom-up skill learning. Submitted.

[] A.M. Turing, (1950). Computing Machinery and Intelligence. *Mind*. Vol.LIX, No.236.

[] M. Turvey, (1992). Affordances and prospective control: An outline of an ontology. *Ecological Psychology*, 4, 173-187.

[] P. Utgoff (1989). Incremental induction of decision trees. *Machine Learning*. Vol.4, 161-186.

[] F. Varela, E. Thompson, and E. Rosch, (1993). *The Embodied Mind*. MIT Press, Cambridge, MA.

[] A. Vera and H. Simon, (1993). Situated action: A symbolic interpretation. *Cognitive Science*. Vol.17, pp.7-48.

[] L. Vygotsky, (1962). *Thought and Language*. MIT Press, Cambridge, MA.

[] D. Waltz, (1990). Eight principles for building an intelligent robot. In: S. Wilson and J. Meyer (eds.) *SAB-90: Simulations of Animal Behavior*. MIT Press, Cambridge, MA.

[] C. Watkins, (1989). *Learning with Delayed Rewards*. Ph.D Thesis, Cambridge University, Cambridge, UK.

[] G. Widmer and M. Kubat, (1996). Learning in the presence of concept drift and hidden context. *Machine Learning*. Vol.23, No.1

[] D. Willingham, M. Nissen, and P. Bullemer, (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 15, 1047-1060.

[] J. Wertsch, (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Harvard University Press, Cambridge, MA.

[] T. Winograd and F. Flores, (1987). *Understanding Computers and Cognition*. Addison-Wesley, Reading, MA.

[] J. Zhang and D. Norman, (1994). Representations in distributed cognitive tasks. *Cognitive Science*. Vol.18, pp.87-122.