

Semiotic Symbols and the Missing Theory of Thinking

Robert Clowes

Centre for Research in Cognitive Science,

University of Sussex

Abstract

This paper compares the nascent theory of the ‘semiotic symbol’ in cognitive science with its computational relative. It finds that the semiotic symbol as it is understood in recent practical and theoretical work does not have the resources to explain the role of symbols in cognition. In light of this argument, an alternative model of symbol internalisation, based on Vygotsky, is put forward which goes further in showing how symbols can go from playing intersubjective communicative roles to intrasubjective cognitive ones. Such a formalisation restores the symbol’s cognitive and communicative dimensions to their proper roles.

Two Kinds of Symbol Systems

Neuroscientist and author of *The Symbolic Species* (1997), Terrence Deacon, has argued that “there is probably no term in cognitive science more troublesome than the word ‘symbol’ ” (2003, p. 117).¹ The problem, according to Deacon, hinges on two notions of the symbol that have been appropriated by different branches of the academy. The first notion is the computational (syntactic) account of symbols known from the Physical Symbol System Hypothesis (PPSH), and the second is the use of symbols as a means of understanding signification and referential systems, especially in language and other forms of communication. This second *semiotic* appreciation of the symbol is based on the analysis of systems of signs, and, in its modern forms, was

developed in two different ways by the French linguist Ferdinand de Saussure and the American pragmatist philosopher Charles Sanders Peirce.

Deacon points out that while the two notions derive from the same intellectual tradition, their development has been such that there is today a deep rift in the conceptual schemes built around them. It may even be that the two notions are now, owing to their separate development, deeply incommensurable. Recent work in cognitive science has, however, suggested that the two notions may be combined or fused. This article will assess the present condition and future possibilities of such a fusion.

The first part of this paper introduces the two theoretical approaches to symbol systems. It then explores the latter *semiotic* approach to symbols in order to investigate what resources it has to explain the sorts of cognitive processes traditionally addressed by symbolically-minded cognitive science. Finding this account lacking, it then sketches a theoretical approach to how symbols developmentally reshape minds. The paper concludes by drawing attention to some future research directions for this vital but neglected area in cognitive science.

Physical Symbol Systems and Cognition

The traditional cognitive science notion of what a symbol is, which we might call the *computational symbol*, is found in similar forms throughout the cognitive sciences² (especially linguistics, philosophy and psychology) and, as Deacon argues, also in mathematics. This notion of the symbol has at its heart the idea that it can support syntactically-based computational operations such as copying, deleting, substituting, and combining.

All of these operations happen according to formal rules with no regard, at base level, for semantics. This idea developed more or less simultaneously across several

disciplines and the development of this *symbolic paradigm* offered unique resources to those thinking about cognition and promised to give an account of mind in purely formal terms. Indeed, it once seemed at least plausible that a complete account of cognition at the psychological level could be given in terms of a formal treatment of systems of symbols, their instantiation in physical processes and their manipulation according to systems of rules.

In the Good Old Fashioned Artificial Intelligence (GOFAI) approach to cognition, coined by Haugeland (1985), the symbol reigned supreme by promising to knit together the worlds of reasoning and representation. With a theory of cognitive architecture organised around the symbol, it was hoped that a fully materialist account of the mind could be spelt out. Such a theory could make minds non-mysterious parts of the physical universe, and according to some, make psychology respectable.

The central reference point for the artificial intelligence understanding of a symbol is Newell and Simon's (1972) idea of a Physical Symbol System (PSS). According to Newell and Simon, the PSS hypothesis not only gave an account of how computational systems can solve well specified problems according to the purely syntactic manipulations of tokens, but it laid out the necessary and sufficient conditions for being an intelligent agent in terms of computational architecture. A PSS can be specified, according to Harnad, in terms of eight conditions. It must have (quoting Harnad):

1. a set of arbitrary "physical tokens" scratches on paper, holes on a tape, events in a digital computer, etc. that are
2. manipulated on the basis of "explicit rules" that are
3. likewise physical tokens and strings of tokens. The rule-governed symbol-token manipulation is based

4. purely on the shape of the symbol tokens (not their “meaning”), i.e., it is purely syntactic, and consists of
5. “rulefully combining” and recombining symbol tokens. There are
6. primitive atomic symbol tokens and
7. composite symbol-token strings. The entire system and all its parts -- the atomic tokens, the composite tokens, the syntactic manipulations both actual and possible and the rules -- are all
8. “semantically interpretable”: The syntax can be systematically assigned a meaning, e.g., as standing for objects, as describing states of affairs (Harnad, 1990, p. 336).³

Fodor (1975; 1987) championed the theoretical justification of a symbol processing system as a theory of thinking principally to specify his computational Representational Theory of Mind (RTM). According to Fodor, a properly constituted theory of the role of symbols in cognition, or in his terminology, *Language of Thought* (or mentalese), is the foundation of a theory of thinking. He argued that tokens in the language of thought were processed in a syntactic matter that is mindful only of their ‘shape’ or formal properties. A further role of the symbol in the RTM was to bind together two apparently very different types of property: the truth-preserving powers of reasoning and the intentional world referring nature of thought.

Mental states were understood as relations to these physical symbols, and mental symbols were thought to have intrinsic representational powers, at least when embedded in the right sort of architecture. They also explained what we really mean when attributing propositional attitudes to other agents, such as in “Jones hopes that X,” or “Mary believes that Y.” Such propositional attitudes were essentially relations to mental symbols. So Jones was related by however his hope mechanisms are

instantiated to the symbols encoding the Proposition X, and Mary by however her hope mechanisms are instantiated to symbols encoding Proposition Y. Symbols allowed us to explain both how reasoning happened and what mental states are.

An important implication was that mental states could thus be attributed to content, and play out inferential episodes of thought in a rational way that – according to Fodor – saved the central assumptions of folk psychology. Their representational powers accrued because of the powers of these internal symbols. As Fodor (1975) argued, “there is no internal representation without an internal language” (p. 55).⁴ As Fodor makes clear, one of the chief benefits of this approach is that it reduces the problem of semantics to formal operations (Fodor, 2003).

Yet Fodor’s notion of what it meant to have an internal language has proved profoundly unsatisfactory. As Fodor has himself admitted, there is little hope that the standard computational theory of symbols or anything much like it is going to explain the sorts of domain general cognition which the human mind seems to support so comprehensively (Fodor, 2000). And if it cannot give a good account of why human beings are rational, it becomes difficult to see the advantages of such a view.⁵ Moreover, recent theories of cognition have made much of how the manipulation of symbols seems to be neither necessary nor sufficient for many properly cognitive episodes (Clark, 1997; Rowlands, 1999) and that much cognition is better conceived of as an entirely non-representational and situated engagement with the world (Brooks, 1991; Dreyfus, 2002). Such embodiment-based or *enactivist* approaches are the major challenger to cognitivism and have promised to give an account of cognition that makes no reference to symbols. Some cognitive scientists who would see themselves within a broadly embodimentalist tradition have, however, now started

to question whether cognitive science could really do without the concept of the symbol.

Semiotic Symbol Systems

A series of theorists (Cangelosi, 2001; Deacon, 1997; Steels, 1999; Vogt, 2003) have recently adopted an alternative approach for understanding symbols. This has arisen in part as a way of dealing with some of the problems faced by the traditional computational symbol.⁶ Its primary focus is to treat the representational abilities, first and foremost, of natural language and other derived and related *semiotic* systems and, in doing so, it is hoped, to show how a revived notion of symbol can once again play a central role in cognitive science. The theoretical background for the *semiotic* symbol which is most often invoked is taken from the work of the pragmatist philosopher Charles Sanders Peirce.

Peirce (1897) developed a formal theory of the sign in respect to its representational capacities, thus “a sign, or representamen, is something which stands to somebody for something in some respect or capacity” (Peirce, 1955, p. 99). He typologised the different kinds of reference that could be established in communication systems, which can be boiled down to the following:

Iconic – A relation of similarity or resemblance, so in the standard example a photograph represents something by virtue of looking like it. According to Deacon’s gloss these simple relations can be understood as an over-generalisation of a learning mechanism.

Indexical – a relation of one-to-one reference or mapping, without similarity between indicator and referent. Sinha (1988) emphasizes how such indexical relations typically rely on causal connections, such as smoke indicating fire. The vervet monkey call-system can be regarded as a paradigmatic “natural”

instance of an indexical communication system. (See Cheney & Seyfarth, 1992 for a detailed discussion of the vervet call system.)

Symbolic – A perceived relation between a sign, a referent and a concept or meaning mediated by conventionality (as explicated below). Because symbols do not generally stand in a one-to-one relationship with objects, an element of interpretation is always required.

The interpretational element of symbolic is a complex issue. Sinha (2004, pp. 223-224) argues:

The conventionality of a true symbol rests on the shared understanding by the communicating participants that the symbol is a token *representing* some referential class, and that the *particular* token represents a *particular* (aspect of) a shared situational context, and, ultimately, a shared universe of discourse. Conventional symbol systems are therefore *grounded* in an *intersubjective* meaning field in which speakers *represent*, through symbolic action, some segment or aspect of reality for hearers. This representational function is unique to symbolization, and is precisely what distinguishes a symbol from a signal. A signal can be regarded as a (possibly coded) *instruction to behave* in a certain way. A symbol, on the other hand directs and guides, not the behaviour of the organism(s) receiving the signal, but their *understanding* (construal) or (minimally) their *attention*, with respect to a shared referential situation.

Sinha emphasises these complex interpretative aspects of the symbol to a much greater extent than Deacon and many other current cognitive-semiotic theorists who emphasise merely the proper instantiation of the internal triadic relationship of a symbol (also derived from Peirce's theory of signs). These allow us to decompose a

symbol into its component relationships which can be most simply explained in terms of (Ogden & Richards, 1923) often reproduced semiotic triangle (Figure 1). According to this view, the symbol can be schematically reduced to a set of triadic relationships among three elements: the representamen (or sign-vehicle), the object and the interpretant.⁷

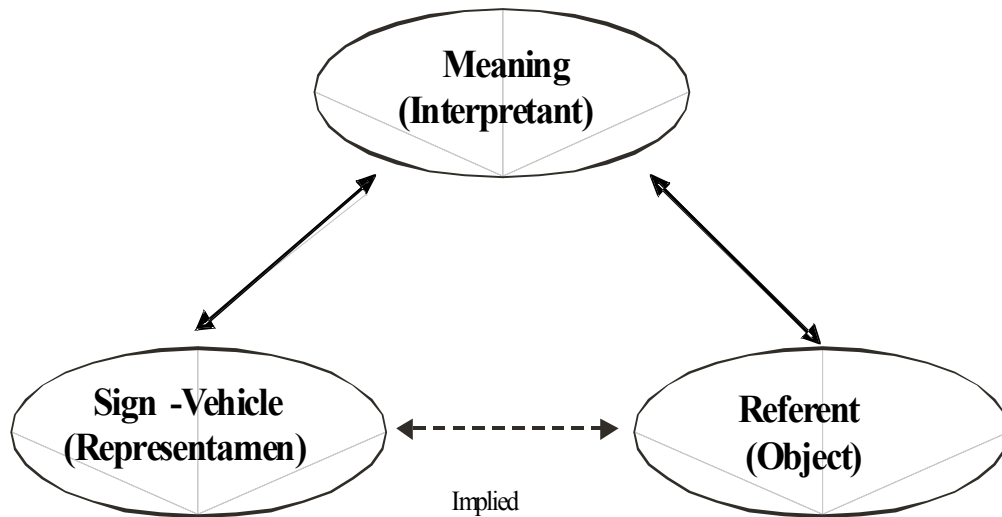


Figure 1. A semiotic triangle after (Ogden & Richards, 1923, p. 11).

Deacon's theoretical innovation was to give meaning to the semiotic symbol in terms of the implicit cognitive architecture that would be necessary to interpret the abstract relationships embedded in systems of properly symbolic signs. Whether a particular communication system is to be accorded the designation of symbolic is, however, a controversial matter. For the purposes of this discussion, I will define *minimal symbol users* as those who meet Deacon's criterion of having a cognitive architecture that can support the interpretation of conventional signs. Such minimal-symbol-users might be capable of only very impoverished symbol interpretation, according to Sinha's criterion. Nevertheless, minimal symbol users interpret signs which are embedded in a system of relationally-defined symbols. Such minimal symbol users can be found in

simulation work such as the artificial agents found in Steels and Kaplan (1999) and Cangelosi, Greco and Harnad (2000).

It remains controversial, however, whether such minimal symbolic capabilities should be linked to the cognitive powers that traditional physical symbol systems are supposed to support. While the mechanics of semiotics might be of value in analysing communicative relationships, why should they be of any use in analysing cognition? In fact, attempts to work out the material basis to the formal relationships specified in Peirce's triad have formed the theoretical underpinning of some important recent projects which have attempted to explain symbol grounding and in doing so have gestured toward explaining some of the novel powers of human thought. While there are many such recent accounts, here I will focus on one project, *The Adaptive Language Games* (or ALG) project developed by Steels and his collaborators (Steels, 1999).

Can Semiotic Symbols Play Cognitive Roles?

Paul Vogt's (2002) *The Physical Symbol Grounding Problem* is the most sustained attempt to show that the ALG framework can show not only how symbols are grounded but also why they should still be regarded as a central concept in cognitive science. In this article Vogt implies that the semiotic symbol approach can solve the problems of symbolist cognitive science, or at least reduce them to "technical" problems by showing how symbols are grounded in communication.

Vogt's approach rests on solving the symbol grounding problem by proposing a rapprochement between embodied cognitive science and some elements of traditional cognitivism. He argues that symbolic structures can be used within the paradigm of embodied cognitive science by adopting an alternative definition of a symbol. In this

alternative definition, the symbol may be viewed as a structural coupling between an agent's sensorimotor activations and its environment.

In Vogt's (2002) paper a robotic experiment is presented in which mobile robots develop a 'symbolic' structure from scratch by engaging in a series of language games. Through the language games, robots construct the means to refer to objects with a remarkable degree of success. Although the underlying meanings (interpretants) of a symbol may vary in different particular language games, agents eventually converge on a system of expressive forms (sign-vehicles) that allows them to pick out referents. That is, the community of agents converge on the same expressive means through communicational episodes, and these episodes in turn structure the agent's internal categorisation of the objects they encounter. The dynamics of the game allows a coherent system of semiotic symbols, in the minimal sense described above, to be developed. This is the basic (yet impressive) result that has been explored from the ALG perspective in a series of papers (Steels & Belpaeme, 2005; Steels & Kaplan, 1999). However, what is interesting for us here is whether these results bear on the question of the role of symbols in thinking.

Vogt develops the basic approach in a discussion of Brooks' earlier work on intelligence without representation (Brooks, 1991). Questioning Brooks' anti-symbolic stance, he asks rhetorically:

But is it true? Are symbols no longer necessary? Indeed much can be explained without using symbolic descriptions, but most of these explanations only dealt with low-level reactive behaviours such as obstacle avoidance, phototaxis, simple forms of categorization and the like (Vogt, 2002, p. 430).

Several theorists (Clark & Grush, 1999; Clark & Toribio, 1994) have raised similar questions. Vogt's preferred solution to the problem comes from a re-interpretation of the symbol along embodimentalist lines. He argues that to overcome the symbol grounding problem, the symbol system has to be embodied and situated. Brooks' *physical ground hypothesis* states "that intelligence should be grounded in the interaction between a physical agent and its environment. Furthermore, according to this hypothesis, symbolic representations are no longer necessary. Intelligent behaviour can be established by parallel operating sensorimotor couplings" (2002, p. 432). Moreover, the way to accommodate symbols in the new situated-embodied perspective is to view them as structural couplings, using Maturana and Varela's (1970) concept. Such an approach is perhaps a reasonable theoretical direction from which one might attempt to subsume symbols into the embodied systems perspective, but it does beg the question of exactly what type of structural coupling they are.

Vogt argues that "when symbols should be necessary to describe cognition, they should be defined as structural couplings connecting objects to their categories based on their sensorimotor projections" (2002, p. 432). This definition, Vogt notes, echoes Peirce's view. Vogt goes on to present something like a standard theory of internal representation with an embodimentalist twist:

Each interaction between an agent and a referent can activate its past experiences bringing forth a new experience. The way these bodily experiences are represented and memorized form the internal representation of the meaning. The actual interaction between an agent and a 'referent' defines the functional relation (2002, p. 434).

But it seems the "symbols" so established are just associations between internal sense and external reference and are constituted simply by establishing the right sort of

association. Such an associationist refiguring of the symbol, however, gives us no way of understanding the difference between the symbolic mode of structural coupling and any other type of structural coupling, and raises the suspicion that what is going on is the formal re-defining of an association as a symbol.⁸

The Missing Theory of Thinking

The ALG approach promises a unification of several dimensions of cognitive science theory. It holds the possibility of providing a mechanistic account of a series of seemingly mysterious processes: how languages are born, how they are maintained, how agents can coordinate categories, and how cultural categories can come into being and be shared across generations.

At first sight it would seem that proposing answers to these questions should open doors to understanding the cognitive role of language. Nevertheless, in examining the ALG approach, it is clear that “symbols” generated in this way cannot be shown, in a straightforward way, to generate traditional cognitive properties. The ALG approach is typically constituted to insulate language from forms of cognitive activity other than categorisation. Despite its argument that language is central to our cognitive adaptivity, the ALG approach actually fixes everything other than the content of the categories in the agent’s architecture.

What is lacking is any sense of how such semiotic symbols play a role in cognitive episodes beyond the picking out of referents in scenes; this is the main task for the agents in all ALG-type experiments. What appears to be missing is the sense that symbols play any role in inferencing or organising non-linguistic behaviour. The worry is that semiotic symbols have come unmoored from cognitive symbols and that therefore our theorising about communicative capacities has come unmoored from

our theorising about cognition. But wasn't this link precisely what the semiotic symbol approach was supposed to theorise?

Perhaps there is much greater indebtedness than would first seem to be in the ALG approach to the GOFAI framework. Although there is no explicit defence of the idea in Steels' work, there is still a ghost of GOFAI in the assumption that in grounding symbols – via the components of the ALG – we can show that symbols also support other cognitive properties. While the ALGs provide architecture for linguistically grounding categories, they make no mention of how such architecture can help an agent to perform other cognitive work. This work seems to be silent on the question of how a symbol system, language system, system of external representations, or system of tools can play a role in reorganising underlying cognitive activity. (For a recent review of the importance of the consideration of this relationship, see Clark, 2006a)

Steels does acknowledge this problem in an article in which he states that the ALGs tell “only the first part of the story. What we still need to show is how these external representations may lead to the significant bootstrapping effect that we see in human development, where representations (drawings, language, pretend play) are a primary motor of cognitive development” (Steels, 2003, p. 14). This is just right but the central problem remains of how to theorise this process.

The semiotic view of the symbol offers a way in which the symbol grounding problem can be solved by offering a materialist explanation of how the dimensions of signifier and signified, or alternatively Peirce's triad of representamen, interpretant and object, could come into relation. It does this by spelling out what it is, at least minimally, for an agent to entertain a symbol and then showing how this can be cashed out in agent-based simulations. But unless some account of how cognitive

architecture can emerge from its ability to interpret symbols is given, a theory of semiotic symbols will never be a serious challenger to GOFAI. The danger of declaring that the symbol grounding problem has been reduced to a technical problem is that it blinds us to the question of the role of semiotic symbols in cognition.

For rationalist and computational accounts of symbol systems, the role of the symbol is to allow inferencing; in essence, an idealisation of thinking shorn of its roots in the ongoing activity of the agent. But if semiotic symbols are held to play inferential or any other type of cognitive roles there must be some theory of how this happens. There appears to be a hidden assumption in Vogt's work on semiotic symbols to the effect that if it can be shown that symbols are grounded, then it can be shown that symbols support truly cognitive properties. But this does not follow. A theory of reference and signification is not a theory of inferencing.

If semiotic symbols are to factor into our accounts of cognition, they face a problem which is every bit as grave as the symbol grounding problem. Although this problem has attracted much less attention than the symbol grounding problem has, it might be dubbed the new problem of symbols. *How do semiotic symbols come to play a role in thinking?*

In what follows I will schematically develop what is needed. This explication will refer to some cognitive modelling work previously reported in Clowes and Morse (2005) that allows us to elaborate on the unique role of symbols in cognition, but it requires attention to how symbols are taken up to do cognitive work. The approach is based on Vygotsky's notion of *semiotic internalisation*.

Restructuring Cognitive Architecture through Symbol Internalisation

Here I present a hypothesis of how developing systems can restructure themselves through the internalisation of symbols. This hypothesis both helps us make sense of

some previously reported simulation-based experiments (Clowes & Morse, 2005) and shows how this work may be linked to Vygotsky's theory of the establishment of higher-cognitive functions (Vygotsky, 1997). Our simulation-based experiments contained agents, embedded in a dynamic environment in which objects could be moved around. Agents were evolved to move objects to target locations within the environment in line with signalled instructions. The neural architecture of the agents was such that they could adapt to re-trigger their own signal reception mechanisms. In these simulations agents came to re-use this re-triggering mechanism to control their own ongoing activity and perform self-regulative functions. Below I argue that this kind of self-regulative function explain the proper cognitive character of the symbol.

According to the following analysis, I argue that we can schematise the three stages of reorganisation that an agent must go through as it internalises symbols:

1. Completing a symbolically initiated action
2. Stabilising activity with symbols
3. Establishing activity regulation with symbols

Completing a Symbolically Initiated Action

If we assume that children are not born knowing what symbols are or how to use them,⁹ we have to assume that they learn about symbols in action. Such a hypothesis requires an *outside-in* understanding of the trajectory of symbols and their role in the regulation and production of behaviour.

However, we certainly should not assume that just because a child can respond appropriately to the use of a symbol that he or she has a fully developed command of symbol interpretation. Vygotsky's colleague Luria (1961) described a mother and child who were playing a game. The mother asks the child, "Where is Lenin?", and the child, having played the game before, points to a painting of Lenin on the wall.

Next, the painting of Lenin is moved and the mother asks again “where is Lenin?” The child points at the place where the picture had hung.

The message should be clear. It is possible for the developing child to have an incredibly imperfect grasp of the thoughts and activities that adults can structure around symbols, and yet no understanding of the deeper interpretative relationships at work. Yet the child is immersed in a world of symbols and some of the most basic interactions with which the child will learn about the world are, from the very first moment, symbolically structured.

In our simulations, agents are confronted with a similar task: to complete an activity sequence which was initiated symbolically from without. As just discussed, this does not mean that the agent needs to understand the meaning of the symbol, if by “understand” we mean some high-level conceptual ability. Our simple simulated agents have no such capacities. However, within acceptable criteria, the agent can interpret the symbol in accordance with the expectations of the speech community from which the symbol has been introduced. The first behavioural regime the agent must go through is the establishment of a behaviour, or group of behaviours (or cognitions) with respect to an instruction, or other form of speech act. We call this the stage of *minimal symbol interpretation*.

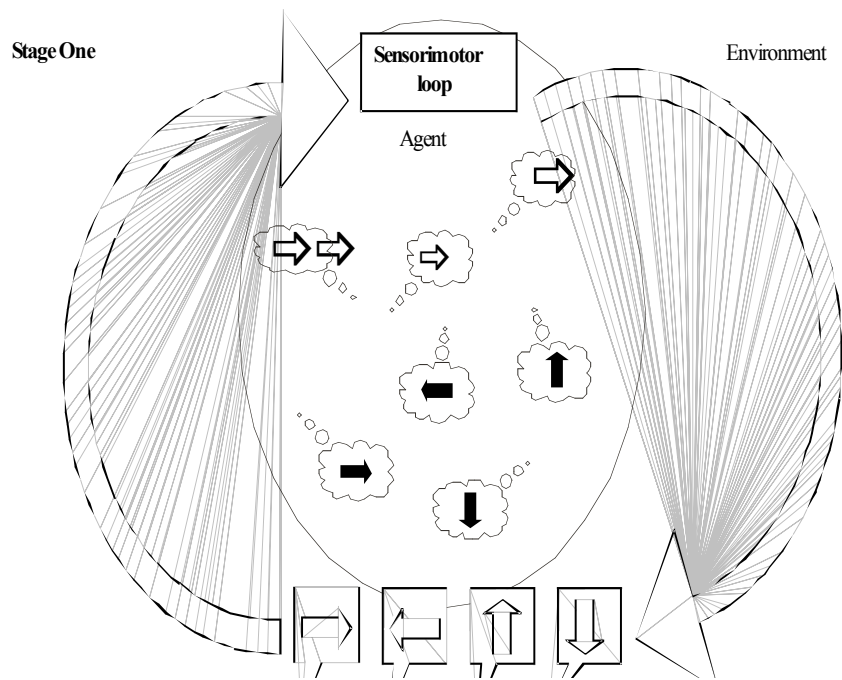


Figure 2, Completing a symbolically initiated action (minimal symbol interpretation).

Figure 2 shows the agent embedded in a series of ongoing interactions (the large curving arrows). The arrows form a rough circle show that these agents are engaged in a continual dynamic interaction with their environment.

In the diagram, some of these interactions are singled out for special attention. These are the externally generated ‘words’ that the agents must interpret, and are shown as square speech bubbles with arrows at the bottom of the diagram. To operate as symbols even in the most minimal sense, these words must be interpreted. Interpretation, at least in this base level language game requires an action. The diagram depicts the neural aspects of these interpretations as thought bubbles, but this is somewhat misleading. Interpretations here really consist of activities, for example, moving objects around in the world. These dynamic interactions with the environment accomplish much of the information processing, or *cognitive* work, of an active

perception system, which need not be interpreted as relying on exhaustive representational systems (Clark, 1997).

This schematic account of the first stage of symbol internalisation merely requires that an agent be able to complete, in the required way, a symbolically-initiated action. One might object that the kind of closely coupled action systems that Luria observed between child and mother could just as well be described as a loosely coupled action system with no real initiator. We could simply consider the mother and child as continually completing, regulating and adapting to each other's activities as in an intimate dance (Cowley, 2007). Yet while the developing relationship between mother and child is built on a whole series of such interactions, the mother initiates their *symbolic character* (as Cowley also recognises). The earliest *symbolically-initiated* actions are when the child completes the mother's action and are thus, as Vygotsky argued, outside-in. It is because symbols first appear for children in such intimate encounters and are only later taken-up by the developing child to structure its own activities that we should understand the process of symbolic development as one of *internalisation*.

Stabilising Activity with Symbols

Words are generally embedded in a series of affective regulation systems that support the construction of activities (Cowley, 2007; Trevarthen, 1991) and are often used to initiate activities from the outside. Typically, these sorts of affective interactions between mother and child provide a series of tacit supports and scaffolds to the child's developing activity system (Bruner, 1983). A whole series of largely unconscious mechanisms seems to be at work in establishing some very basic social psychological functions, such as triadic interactions (Leavens, 2006). In addition, of

course, there are some very conscious interactions, as the mother seeks to engage her child in aspects of the surrounding world.

Yet in the midst of these developing interaction systems, the child also faces a problem. If symbolic regulation is to play a role in structuring the child's own autonomous activities, *symbols must be wrested from their public source and appropriated for self-directed activity*. Appropriation of symbols requires performance without some of those social scaffolds. Accomplishing this task seems to require the development of internal mechanisms which can take over the role of some of these supports.

Vygotsky discussed similar problems in his writings on the question of the differentiation of functions of egocentric speech (Vygotsky, 1986). As egocentric speech develops towards properly self-regulative internal-speech, there is developmental evidence that the child has difficulty in wresting this speech from its social source. Vygotsky writes, "in the process of growth the child's social speech, which is multifunctional, develops in accordance with the principle of the differentiation of separate functions, and at a certain age it is quite sharply differentiated into egocentric and communicative speech" (cited in Wertsch, 1985, p. 117). In the first stages of the development of using self-directed speech for control, as Wertsch notes, Vygotsky "reasoned that one should find a lack of differentiation or even thorough confusion between social and egocentric speech in young children's verbal behaviour" (Wertsch, 1985, p. 118).

Vygotsky empirically tested this hypothesis of progressive differentiation with three experiments, each of which was designed to test the child's ability to differentiate social from egocentric speech in action. Egocentric speech, according to Vygotsky, is by definition involved in self-control functions. Yet social contact was

found to regulate the production of egocentric speech in a variety of experimental conditions. Children use much less egocentric speech in situations where they have less chance of being understood by others. Vygotsky explained this as an initial difficulty with differentiating social from self-organisational functions.

Vygotsky frames the problem in terms of how a child has to learn to use self-directed speech in the absence of adults or other children. This is difficult because speech must be turned from its social function. The problem of gaining control of these minimally-symbolic interpretation systems is also a problem for the agents in our simulations (Clowes & Morse, 2005). This problem confronts the agents in a number of forms.

First, there is simply the problem of developing and using a self-directed loop. In our simulations, agents have the capacity to trigger themselves because of their re-entrant architecture. However they initially switch this off because self-triggering interferes with their capacity to interpret signals generated from outside. Self-generated signals may be mistaken for external ones and upset the developing activity pattern. To take advantage of self-generated signals, agents must learn to differentiate those that are self-generated from those that emanate from outside. The problem of appropriating symbols to self-control manifests itself in the agent simulations. As reported in Clowes and Morse, early generation agents tend to turn off their internal loops until initial control regimes are stabilised. When they turn the loop back on at a more advanced stage, there is some initial drop in performance.

This is essentially a problem of self-organisation. The agent has developed interpretations, responses and structural couplings that are cued by “good” information, that is, information that the agent needs to achieve its goals. The advent of self-directed signals requires the restructuring of an agent’s interpretation

mechanisms as these can destabilise behaviour. Such potential for destabilisation predicts a U-shaped curve in developmental episodes when self-signalling becomes involved in self-regulation activity. However, regimes of self-stimulation begin to make new modes of self-regulating activity possible. At this point, the agent's self-stimulating use of its own interpretation mechanisms reflects more of a potential for activity than an actual new organisational regime of activity.

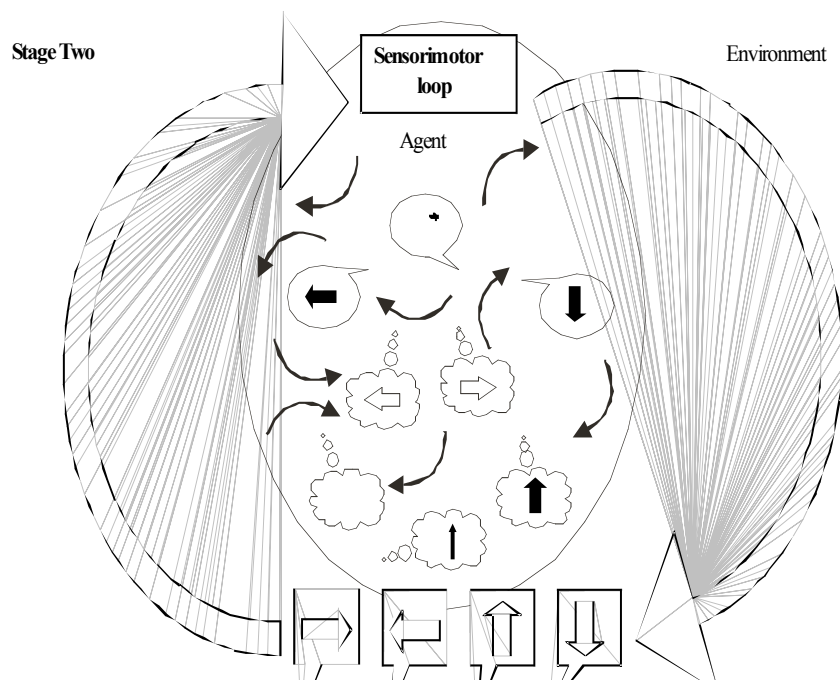


Figure 3. Learning to produce a symbol as a cue to action.

Figure 3 represents this transitional stage of development by showing a series of internally-generated speech bubbles that can trigger interpretation processes, and which can in turn trigger the production of internally-generated “speech,” (hence the internal partial loops shown by the small curved arrows). This stage of development implies a second type of cognitive architecture that develops as the agent starts to use symbols to stabilise its activities. This phase of activity re-ordering can begin once some symbol interpretation systems are in place. The establishment of action-based

interpretation systems form a new platform on which the agent constructs new modes of action and self-regulation.

This stage of development is an unstable and transitional point as most of the agent's self-directed signals can be regarded as noise, but noise that has the potential to become a new kind of self-directed activity. In this phase the agent is faced with both problems and opportunities. As self-generated auto-stimulation loops become stabilised, the agent has the possibility of organising its activities according to new means of control that are established at a higher, semiotically generated, order of abstraction.

Establishing Activity Regulation with Symbols

In this third stage, agent organisation has gone beyond the simple need to establish when an “utterance” comes from outside and when it is produced internally. It has therefore differentiated for itself *in practice* inside and outside. This differentiation allows new types of functional differentiation to take place. Now the agent is in a position to capitalise on these newly developed internally-directed speech circuit loops and to develop entirely new modes of activity. To do this, it needs to establish when auto-stimulation with words is useful and when it is unnecessary. In a sense, it needs to establish mastery of the sensorimotor contingencies of its own activity system.¹⁰ As Vygotsky pointed out, this sort of development requires the progressive differentiation of functional systems that respond to externally-generated activity from those that respond to internally-generated activity.

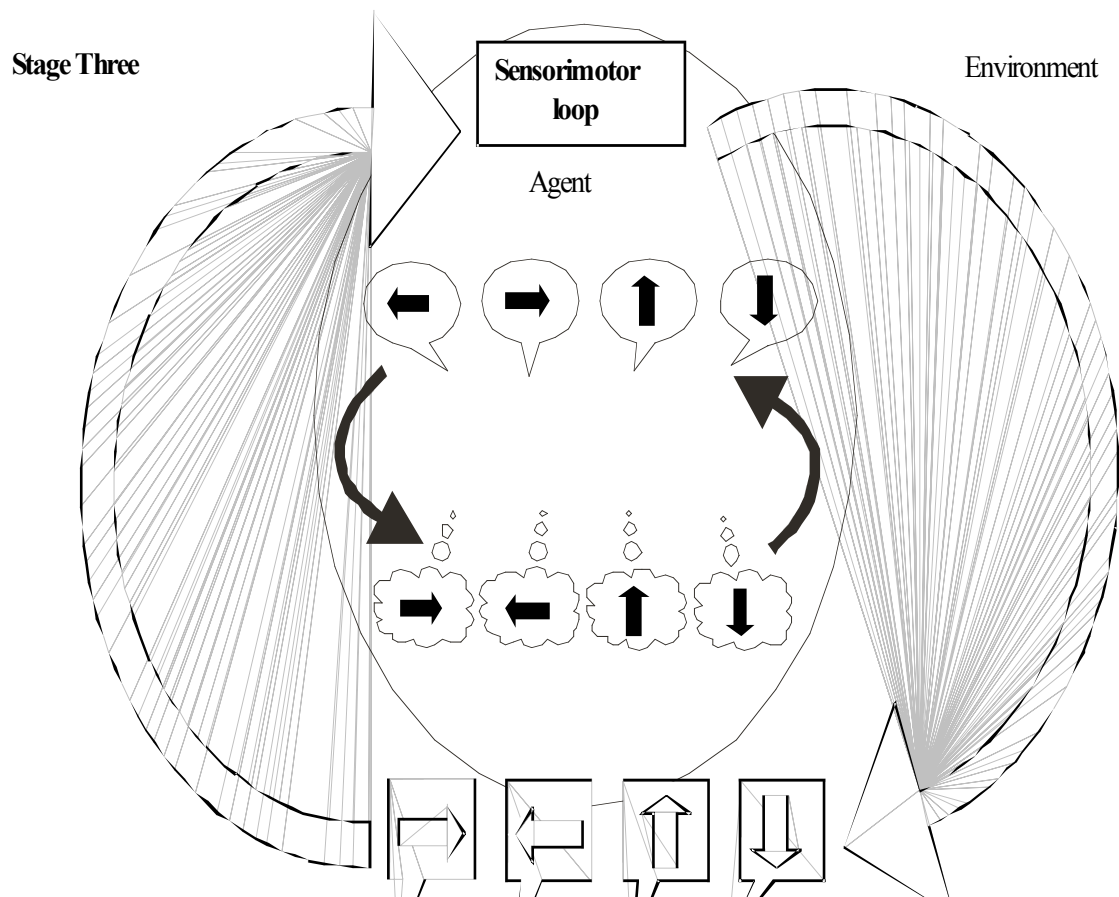


Figure 4 . Establishing activity regulation with symbols.

In Figure 4, this is represented by the development of a new, functionally differentiated and *internal* activity loop. Unlike in the previous control regime, internally-generated loops do not simply capitalise on externally-generated and supported activity systems, but they develop new activity systems. The functional organisation of the activity of an individual agent has a logic that is not simply a recapitulation of the logic of the group. Public systems of representation produced socially are thereby turned to the agent's own ends. This point of development could be regarded as the point of completion of symbol internalisation, for the agent has now built a new mode of symbolically-mediated self-regulation that is essential to its ongoing activity.

Towards an Understanding of Semiosis in Cognition

The theoretical model presented here and the experimental results presented in Clowes and Morse (2005) give us the beginnings of an account of how the internalisation of symbols come to reshape neural-dynamics. In contrast to other accounts of semiotic symbols systems, these models illuminate the neglected cognitive side of the semiotic symbol.

At a technical level, these models indicate one manner in which using external symbols can reorganise the basic mechanics of regulating activity in a minimal cognitive model. (Work is ongoing to understand this process in more detail.) At a more abstract level, they give us a sense of how activities and the structural couplings between agents and their environments can be stabilised around the concrete anchors made available by semiotic systems. Nevertheless, there is much work to do. This account represents only the beginnings of an understanding of how symbol internalisation reorganises cognition in human beings. To deepen this understanding we need to tackle the following questions:

1. How do words, and the external social representational systems in which they are embedded make available the contingencies through which agents restructure themselves?
2. What forces are at play in the internal dynamics of agents such that they can appropriate these structures?
3. How do the external and internal systems interact in the ongoing restructuring of agents?

The outside-in model forces us to treat explicitly how the representational structure of language allows an agent to shape its own cognitive architecture. I have argued for the need to understand this process through a sequence of functional changes. The first is how scaffolding (Bruner, 1983) gives way to semiotically-mediated joint control (Cowley, 2007) and how this in turn gives way to semiotically-mediated self-control (as Vygotsky emphasized). This process has profound implications for the attentional systems of the developing child, its sense of self and its agency; to understand this we need research on the unfolding functional changes that underpin internalisation.

Acknowledgments

The simulation work described above was carried out in association with Anthony Morse, and Sean Bell originally produced the diagrams. This paper has also benefited from the criticism of a panel of anonymous referees and much helpful advice from the special edition editors to whom I would like to express my thanks.

¹ Whereas this is an accurate summary of today's state of affairs, we should add that the notion of symbol has become vastly more problematic as traditional GOFAI approaches to cognitive science have fallen into disfavour. The symbol has become a problem insofar as the more general research programme has become increasingly problematic.

² The symbols which we find in AI, the philosophy of cognitive science and the generativist tradition in linguistics, which sees the formation of sentences as formal manipulations of syntax, share much in

common. They are not, however, identical notions, and there is a nice discussion of some of the subtle differences between them in Chapter 7 of (Rowlands, 1999).

³ I cited at length from Harnad here because his view seems to be a fairly canonical one about what symbols are in the cognitive science community, even if there is not wide agreement on whether his view on how symbols are grounded is correct.

⁴ A curious upshot of this idea in its classical form is that language is a subsidiary phenomenon only made meaningful when it is translated into an inner language.

⁵ This is, of course, to leave aside the attendant problems of the seemingly unavoidable commitment to conceptual nativism (Fodor, 1998).

⁶ Actually, proposals for incorporating semiotic theory into the understanding of cognitive organisation have some longstanding proponents (Sinha, 1988). Perhaps the recent resurgence of interest in this area can be linked to the development of new techniques for modelling multi-agent systems, some of which are discussed below.

⁷ The original terminology used by Peirce in characterising the triadic relationship was *representamen*, *interpretant* and *object*. Vogt, Steels and his colleagues tend to use the terms *form* (or *word-form*), *meaning* and *referent* (Vogt, 2003). I prefer to refer to the representamen as *sign-vehicle* as this emphasises its role in conveying meaning.

⁸ In fact, if the main idea behind the semiotic conception of symbols is correct, properly the identification of a symbolic relationship would be a difficult task to perform at the level of the individual structural coupling of an agent and its environment. This is because the right sorts of structural couplings are not defined by the individual relationships, but by the *system* of relationships in which they are embedded. Vogt's attempted definition is at the very least missing a crucial feature.

⁹ As a number of theorists have argued, for example, (Clark, 2006b; Cowley, 2005; Sinha, 1988; Vygotsky, 1986).

¹⁰ This use of terminology is a gesture toward the theorisation of active perception developed in (O'Regan & Noë, 2001).

References

- Brooks, R. (1991). Intelligence without Representation. *Artificial Intelligence*(47), 139-160.
- Bruner, J. S. (1983). *Child's Talk*. Oxford: Oxford University Press.
- Cangelosi, A. (2001). Evolution of Communication and Language Using Signals, Symbols, and Words.
- Cangelosi, A., Greco, A., & Harnad, S. (2000). From Robotic Toil to Symbolic Theft: Grounding Transfer from Entry-Level to Higher-Level Categories. *Cognitive Science*, 12(2), 143 - 162.
- Cheney, D. L., & Seyfarth, R. M. (1992). *Précis of How Monkeys see the world*. *Behavioral and Brain Sciences*, 15(1), 135 -182.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: The MIT Press.
- Clark, A. (2006a). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8), 370-374.
- Clark, A. (2006b). Material Symbols. *Philosophical Psychology*, 19(3), 291-307.
- Clark, A., & Grush, R. (1999). Towards a Cognitive Robotics. *Adaptive Behavior*, 7(1), 5-16.
- Clark, A., & Toribio, A. J. (1994). Doing without Representing. *Synthese*, 10, 401-431.
- Clowes, R. W., & Morse, A. (2005). Scaffolding Cognition with Words. In L. Berthouze, F. Kaplan, H. Kozima, Y. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov & C. Balkenius (Eds.), *Proceedings of the 5th*

-
- International Workshop on Epigenetic Robotics*. Nara, Japan: Lund University Cognitive Studies, 123. Lund: LUCS.
- Cowley, S. J. (2007). How infants deal with symbol grounding. *Interaction Studies*
- Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the human brain*: The Penguin Press, Penguin Book Ltd.
- Deacon, T. W. (2003). Universal Grammar and semiotic constraints. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution: The States of the Art*: Oxford University Press.
- Dreyfus, H. L. (2002). Intelligence Without Representation. *Phenomenology and Cognitive Science*, 1(4), 367 - 383.
- Fodor, J. (1975). *The Language of Thought*. New York: MIT Press.
- Fodor, J. (1987). *Psychosemantics*: MIT Press.
- Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. *The 1996 John Locke Lectures*. (Vol. Oxford University Press.).
- Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge MA: MIT Press.
- Fodor, J. (2003, 9 October). More Peanuts: Review of José Luis Bermúdez *Thinking Without Words*. *London Review of Books*.
- Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42, 335-346.
- Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, Massachusetts: Bradford/MIT Press.
- Leavens, D. A. (2006). It takes time and experience to learn how to interpret gaze in mentalistic terms. *Infant and Child Development.*, 9, 187-190.
- Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior*. New York: Pergamon Press.

-
- Maturana, H. R., & Varela, F. J. (1970). *Autopoiesis and Cognition. The Realization of the Living*: Dordrecht.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*: Englewood Cliffs, NJ: Prentice-Hall.
- Ogden, C. K., & Richards, I. A. (1923). *The Meaning of Meaning: A study of the influence of language upon thought and the science of symbolism*. London: Routledge and Kegan Paul, Ltd.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24.
- Peirce, C. S. (1955). *Philosophical Writings of Peirce*. New York: Dover Publications.
- Peirce, C. S. (Ed.). (1897). *Logic as semiotic: The Theory of Signs* (1985 ed.): Indiana University Press.
- Rowlands, M. (1999). *The Body In Mind: Understanding Cognitive Processes*. Cambridge: CUP.
- Sinha, C. (1988). *Language and Representation: a socio-naturalistic approach to human development*: Harvester Wheatsheaf.
- Sinha, C. (2004). The Evolution of Language: From Signals to Symbols to System. In D. Kimrough Oller & U. Griebel (Eds.), *Evolution of Communication Systems: A Comparative Approach* (pp. 217-237). Cambridge MA: MIT Press.
- Steels, L. (1999). *The Talking Heads Experiment: Volume I. Words and Meanings* (Pre-Edition ed.). Antwerpen: Laboratorium.
- Steels, L. (2003). Intelligence with Representation. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 361(1811).

-
- Steels, L., & Belpaeme, T. (2005). Coordinating Perceptually Grounded Categories through Language: A Case Study for Colour. *Behavioral and Brain Sciences*, 28(4), 469--489.
- Steels, L., & Kaplan, F. (1999). Collective learning and semiotic dynamics. In D. Floreano, J. D. Nicoud & F. Mondada (Eds.), *Advances in Artificial Life (ECAL 99), Lecture Notes in Artificial Intelligence* (pp. 679-688). Berlin: Springer-Verlag.
- Trevarthen, C. (1991). The function of emotions in early infant communications and development. In J. Nadel & L. Camaioni (Eds.), *New Perspectives in early infant communication and development*. (pp. 48-81). London: Routledge.
- Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, 3(3), 429-457.
- Vogt, P. (2003). Anchoring of semiotic symbols. *Robotics and Autonomous Systems*, 43(2), 109-120.
- Vygotsky, L. S. (1986). *Thought and Language* (Seventh Printing ed.): MIT Press.
- Vygotsky, L. S. (1997). The History and Development of Higher Psychological Functions. In R. W. Rieber (Ed.), *The Collected works of L. S. Vygotsky. Vol.4*, . New York ; London: Plenum.
- Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind*. Cambridge Mass: Harvard University Press.