The Epigenesis of Meaning in Human Beings, and Possibly in Robots

JORDAN ZLATEV
Department of Linguistics, Lund University, Sweden; Department of Linguistics, Chulalongkorn University, Thailand; E-mail: jordan.z@chula.ac.th, jordan.zlatev@lucs.lu.se

Abstract. This article addresses a classical question: *Can a machine use language meaningfully and if so, how can this be achieved?* The first part of the paper is mainly philosophical. Since meaning implies intentionality on the part of the language user, artificial systems which obviously lack intentionality will be 'meaningless' (*pace* e.g. Dennett). There is, however, no good reason to assume that intentionality is an exclusively biological property (*pace* e.g. Searle) and thus a robot with bodily structures, interaction patterns and development similar to those of human beings would constitute a system possibly capable of meaning – a conjecture supported through a Wittgenstein-inspired thought experiment. The second part of the paper focuses on the empirical and constructive questions. Departing from the principle of *epigenesis* stating that during every state of development new structure arises on the basis of existing structure plus various sorts of interaction, a model of human cognitive and linguistic development is proposed according to which *physical*, *social* and *linguistic* interactions between the individual and the environment have their respective peaks in three consecutive stages of development: *episodic*, *mimetic* and *symbolic*. The transitions between these stages are qualitative, and bear a similarity to the stages in phylogenesis proposed by Donald (1991) and Deacon (1997). Following the principle of epigenetic development, robotogenesis could possibly recapitulate ontogenesis, leading to the emergence of intentionality, consciousness and meaning.

Key words: consciousness, development, embodiment, epigenesis, intentionality, language, meaning, robotics, social interaction, the Turing Test

1. Introduction

Imagine a typical (within a Western middle-class sociocultural context) situation in which a parent – let us make it the father – is playing with his two-year-old child. The two are sitting on the floor, surrounded by various toys. The child is playing with some of these, and sometimes hands one over to Dad. Language enters the play quite naturally. The child may gesture toward a toy train engine that happens to be out of reach and say *train*. Dad then responds both verbally: *Oh, you want the train engine? Here you are!* and non-verbally by passing the requested toy. In receiving it, the child repeats *train-engine, train-engine*, thereby indicating that the adult's slight correction concerning the proper term of reference has not passed unnoticed. Later on, they exchange roles and Dad asks: *Can you please give me that red car?* – indicating through his gaze and gesture a particular Matchbox model. Being in a co-operative mood the child responds by giving daddy the car. *Thank you*, says Dad.

Minds and Machines **11:** 155–195, 2001.

The activity described in the passage above is an example of a *language game* – the influential concept introduced in the work of the late Wittgenstein, at least according to one of its possible interpretations.

> Language games are the forms of language with which a child begins to make use of words ... When we look at the simple forms of language the mental mist which seems to enshroud our ordinary use of language disappears. We see activities, reactions, which are clear-cut and transparent. On the other hand we recognise in these simple processes forms of language not separated by a break from our more complicated ones. We see that we can build up the more complicated forms from the primitive ones by gradually adding new forms. (Wittgenstein, 1969: 17)

The 'mental mist' Wittgenstein refers to consists of notions such as 'communicative intention', 'meaning', 'grammar' and 'understanding' which continue to baffle the cognitive sciences – especially when conceived as exclusively private, mental phenomena. At the same time, in the interaction between parent and child described above, there doesn't seem to be anything particularly mysterious: The *intention* of the child is most often transparent, as any parent would testify, despite the still 'primitive' nature of its speech.[1] The *meaning* of the child's words can quite plausibly be identified with the role they play within the language game. The child's *understanding* (or non-understanding) is also evident from its responses – in the absence of (fairly obvious) reasons not to co-operate, children generally do. Finally, the *grammar* of the speech of the average 2-year-old is still quite simple (cf. Fenson et al., 1994; Bates, in press), but if indeed there is a continuous progression from the 'primitive' to the 'more complicated' expressions in the child's language as conjectured by Wittgenstein, and as empirically documented by Tomasello (1999), then there is no reason to postulate any innate grammar modules in the brain (cf. Chomsky, 1980; Pinker, 1994), which furthermore would be anomalous from the standpoint of evolution (cf. Müller, 1994; Deacon, 1997).

Now imagine that the language game described above is more or less unchanged – with one notable exception: *in the place of the child there is a robot.* The robot can pick up and 'play' with the different toys; it can hand them over to the human 'caregiver' on verbal request and follow his gaze when necessary; it can also point to objects of interest and initiate verbal interactions: *May I have the ball?* In one sentence: The robot has the physical, social and linguistic competence corresponding to a 2–3-year-old child, and is able to participate in adequate for its 'mental age' language games.

In relation to this little thought experiment, we may ask two questions. The first one is mainly conceptual and the second mainly empirical, though their answers are not independent.

- If a robot is able to participate in simple language games as adequately as a child, should we concede true meaning and intelligence to it?

- How would we go about developing a robot which could possibly live up to a positive answer to the first question?

Section 2 will address the first question, and to anticipate, will propose a positive answer: Yes, a robot could in principle be said to possess genuine intelligence and meaning – provided that it exhibits crucial properties of *human* intelligence and meaning. Which are these? Following earlier work (Zlatev, 1997) and a simple thought experiment I will single out the following three: (a) sociocultural situatedness, (b) naturalistic embodiment and (c) epigenetic development. I will argue that such an answer is in important ways different from familiar proposals for 'operational definitions' of intelligence by Turing (1950) and Harnad (1991), and also immune from refutations of these presented by e.g. Searle (1980, 1995) and Dreyfus (1991, 1993).

But since it is not my intention to make this paper purely philosophical (for which, in any case, I lack sufficient qualifications) the major emphasis will be on the second question, addressed in Sections 3 and 4. First, a research program, *Epigenetic Robotics*, aiming at developing a robot according to the principles discussed in Section 2 is briefly presented. In terms of its goals and guiding principles it is shown to be similar to an existing, well-known robotics project – *Cog* (Brooks et al., 1999). However, I will argue that there is a discrepancy between the goals and the actual achievements of the latter, showing that the *Cog* project lacks a conceptual framework of how "higher" cognitive abilities, including intentionality and consciousness, could possibly emerge in a robotic system. Since such abilities are both *presupposed* and *further developed* by meaningful language use, without developing such a framework, the goals of a project such as *Cog* will remain utopian. The ideas presented in Section 4 are intended as a first attempt to formulate such a framework. Section 5 summarizes the main points of the argument.

2. Can a Machine Mean?

2.1. The Turing Test and its Critics

The title of this section is a paraphrase of the familiar question "Can a machine think?" which opens Turing's (1950) classical paper introducing what came to be known as the *Turing Test* (TT) for machine intelligence. The adequacy of the paraphrase lies in the fact that like many others, Turing regarded the ability to use language as the best criterion for intelligence. If a computer program can *simulate* the role of a conversation partner – over a terminal in order to hide some inessential details such as the absence of a body – so well as to make it impossible for its interlocutor to decide whether it is a program or a person, then one should conclude that it can indeed think, according to Turing. Not only did he claim that his thought experiment showed that artificial intelligence was achievable *in principle*, but predicted that in about 50 years – that is *now* – given a five-minute interview on a free subject, an average human interviewer would not have more than 70%

chance of deciding whether he is chatting with a con-specific or with a computer. Given the state-of-the-art of current 'natural language processing' systems, Turing's more specific prediction would be validated only if the intelligence of the 'average questioner' is quite low.

The in-principle part of Turing's argument has not fared much better. In his three-decade long debate with the orthodox AI community Dreyfus (1973, 1991) has successfully argued that in order to be able to engage in even the simplest conversations, a machine would have to possess an enormous amount of *background knowledge*. And it can not be a matter of just 'programming in' this knowledge – as in 'knowledge bases of common sense knowledge' such as that developed within the CYC project of Lenat and Feigenbaum (1991) – because it consists not of 'facts', 'rules', and other such-like propositional representations, but of practical skills, bodily dispositions and cultural practices. These constitute a kind of tacit knowledge, or know-how, which can be obtained only through experience.

But perhaps the strongest, or at least the most influential, criticism of the adequacy of Turing's operational definition of intelligence, and hence of the justification for a literal interpretation of the term Artificial Intelligence, is Searle's well-known "Chinese Room argument" (cf. Searle (1980) for the original and Searle (1995) for a recent formulation). To remind: the argument states that passing the TT can not be a guarantee for *understanding*, since a speaker of one language could in principle implement a computer program (written in a language which he does understand) and according to it *mechanically* perform input-output operations and various internal manipulations on the messages of a totally unfamiliar to him language. What from the outside may count as passing the TT is for the person inside the room meaningless 'symbol-manipulation' – he has no idea either what the words he is saying *mean* (or even if they are words at all) and he can not be said to have any kind of *communicative intention*. The moral is: "syntax is not the same as, nor by itself sufficient for semantics" (Searle, 1995: 61) and since a computer can only operate on syntactically defined 'rules',[2] there is no way it could either mean or understand anything. All meaning and intelligence would be, in the best case, in the eyes of the beholder.

2.2. SOME DEFENDERS OF THE TT

Searle's argument relies on intuitions, and intuitions are (almost) never watertight, but it is nevertheless better than its many published 'rebuttals'. Churchland and Churchland (1990), for example, attack the premise that syntax is not sufficient for semantics (i.e. the moral from the Chinese room thought experiment) claiming that it is analogous to the false premise "forces by themselves are ... not sufficient for luminance" (ibid: 29) which leads to the false conclusion that "electricity and magnetism [which are forces] are not sufficient for light". Just because the premise *seems* right to us does not guarantee that it is right – computation may very well give rise to (what we could call) semantics. The problem with this

counter-argument is that the analogy between light-luminance and mind-meaning is bad: the first are physical properties and clearly reducible to more basic physical properties and physical laws. Meaning on the other hand, implying notions such as conventions and communicative intention, is not (or at least not only) a physical property and it is not at all clear how it could be reduced to physical properties, and even less so to the kind of 'computations' performed in a digital computer.

Another critic of Searle and thus (indirectly) defender of the TT, Dennett (1991) gives a version of the 'systems reply': a program that would be able to pass the TT (even for 5 minutes'?) "would have to be an extraordinately supple, sophisticated, and multilayered system, brimming with "world-knowledge" and meta-knowledge, and meta-meta-knowledge about its own responses, the likely responses of its interlocutor, and much, much more" (ibid: 438). But if the program is so complex, continues Dennett, then our intuitions can no longer tell us that there is no genuine understanding going on – unlike in Searle's simplistic scenario of a man mindlessly manipulating meaningless characters: "Maybe the billions of actions of all those highly structured parts produce genuine understanding in the system after all." (ibid: 438). Here, Dennett appears to be endorsing a kind of who-knows?-maybe-syntax-could-give-semantics 'emergentist' position – though as pointed out above, such a position has its own problems.[3] However, later on he states that what the mindless subroutines of the program are performing is a 'quasi-understanding', which non-AI practitioners are just too unimaginative to see for what it is: "Probably because they haven't learned how to imagine such a system. They just can't imagine how understanding could be a property that emerges from *lots of distributed quasi-understanding in a large system*." (ibid: 439, my emphasis). I frankly admit to share this lack of imagination. In particular: I see no other reason to elevate the subroutines (elsewhere described by Dennett (1991) as simple mindless 'demons' specialized for certain input-output operations) to the status of 'quasi-understanders' than Dennett's desire to mix up causal and intentional processes in his magician's hat and somehow make the latter disappear. Dennett's 'theory of consciousness' seems to add up to the claim that consciousness, in its various manifestations – meaning, introspection, qualia, etc. – is some kind of elaborate self-illusion resulting from "the operation of a "Von-Neumanesque" virtual machine implemented in the parallel architecture of a brain" (ibid: 210). Unsurprisingly, this characterization has failed to satisfy any but the most sympathetic of Dennett's readers.

2.3. ENTER THE ROBOTS

More constructive responses to Searle's argument are those which fall into the "robot reply". One of these belongs to Harnad (1991, 1993) who proposes extending the TT into the "Total Turing Test" (TTT) on the grounds that:

> . . . in the case of the TT, there was more we could ask for empirically, for human behavioral capacity includes a lot more than pen-pal (symbolic) interactions. There is all of our sensorimotor capacity to discriminate, recognize, identify, manipulate and describe the objects, events and states of affairs in the world we live in. Let us call this further behavioral capacity our *robotic* capacity. Passing the TTT would then require indistinguishability in both symbolic and robotic capacity. (Harnad, 1992, 6.2)

Harnad's point is that if we could not distinguish a robot from a human being "by exactly the same criteria we use in judging one another in our ordinary, everyday solution to the other minds problem" – by which Harnad obviously means competence in physical and social interaction, rather than, say, having skin, a bodily odor and so on – we would have no principled basis for doubting that it lacks intelligence such as ours.

This is a fairly persuasive argument, and I agree with Harnad that a machine possessing such abilities would be immune to Searle's argument since its successful operation would require not only blind 'symbol manipulation', but causal and 'indexical' links to the environment. Even if there is a homunculus who is implementing the central processor in this "walking Chinese room", he can not know what is going on in this extended embodied system since he is only a small part of it. But the non-applicability of Searle's argument does *not* in itself show that passing the TTT test, or behavioral 'indistinguishability', is a sufficient condition for accrediting the robot with meaning, which Harnad acknowledges as well.

Let us consider the following Wittgenstein-inspired thought experiment:[4] a person who has lived a normal life in our community dies and in the autopsy it is discovered that there is some kind of a device instead of a brain in his head. Would we on the basis of this decide that we had been fooled all along and that the person was actually a 'brainless' automaton, lacking any real language and meaning? I believe that the answer is: hardly. Notice, however, the following differences in this thought experiment compared to the earlier ones.

First, the person-robot has not participated in a test, of any arbitrary duration, but has engaged in natural everyday activities *within a community for a whole lifetime*. This seems to exclude the possibility that he has been constructed with the explicit purpose of fooling the judges of a TTT test.

Second, the 'death' of the person-robot makes it too late to investigate the causal relationship between his 'hardware' and behavior, and even if we may have some doubts we have no way of substantiating them. Thus we remember what a great guy the person-robot was, how he laughed at parties and helped old ladies cross the street and decide that *whatever* it was that let him have a soul (in a sense), he certainly had one. Otherwise it would simply be unfair toward the deceased. But what if he hadn't died? Then first the doctors at the operation table, then the cognitive scientists and then we, would have wanted to know: OK, if not a brain, what is it that makes him tick? And if we were told that it is a bunch of cogs, or

even a mini supercomputer – both of which are completely dissimilar to our brains – then we would become suspicious that maybe he did not have a real mind after all. And we would want more evidence on how he actually 'worked'. The point is that *intuitions about causality matter for intuitions about personhood* – at any rate in a scientific culture.

Third, a very important source of evidence in deciding whether our neighbor the robot had genuine intentionality or not – once we are made open to the possibility that he might not – is the answer to the question: *how did he acquire* all the physical and social skills necessary for us to think of him as just one of us? If we are told (after our cognitive scientists work on him extensively, either dead or alive) that it was through preprogramming by super-intelligent engineers (or perhaps by aliens) then again we would be disposed to accept the possibility that he was an ingenious automaton after all. On the other hand, if we were assured that he grew up and learned his skills just like us, this suspicion would hardy arise.

2.4. Artificial situated embodiment

What I would like to propose on the basis of this discussion, is that the following three characteristics of human cognition in general, and language use in particular (cf. Zlatev, 1997) constitute criteria for the attribution of intelligence and meaning to an artificial system:

- *sociocultural situatedness*: the ability to engage in acts of communication and participate in social practices and 'language games' within a community;
- *naturalistic embodiment*: the possession of bodily structures giving adequate causal support for the above, e.g. organs of perception and motor activity, systems of motivation, memory and learning; (notice that this implies *structural* similarity between a natural and artificial system, not physical similarity, and absolutely not identity);
- *epigenetic development*: the development of physical, social and linguistic skills along a progression of levels so that level *n*+l competence results from level *n* competence coupled with interaction with the physical and social environment.

Furthermore, the degree of necessary similarity between human and artificial intelligence with respect to these features is an open issue – it is definitely not "indistinguishability" we need to strive after. In sum, I propose the following criterion for giving a positive answer to the question "Can a machine mean?":

> *If an artificial autonomous system (a robot) with bodily structure similar to ours (in the relevant aspects) has become able to participate in social practices (language games) by undergoing an epigenetic process of cognitive development and socialization, then we may attribute true intelligence and meaning to it.*

Let me attempt to answer two likely objections to this proposals – one from the 'Left' and one from the 'Right', so to speak. First, from a reductive materialist position (in unison with behaviorism and the Turing Test) one could question the necessity of epigenetic development. After all, if a preprogrammed robot behaved in the same way as one which had developed through interaction with the physical and social environment – to the extent that the two could not be distinguished – then on what basis can one deny the first one intelligence? Epigenesis may be necessary to discover the 'goal-state' of the robot's 'brain', but once this is found, why not implement this state directly in the future? The more philosophical response to this objection is that the pre-programmed robot will have nothing but (first-order) *derived intentionality* – all its 'representations', 'goals', 'beliefs', etc. would derive their meaning entirely from the intentionality of the engineers who programmed it. The robot itself would be performing whatever it was designed to perform blindly. Even Dennett (1996), well-known for his rejection of intrinsic intentionality appears to acknowledge that there is an important distinction between programming-in and learning through experience in a thought experiment about a milk-buying robot:

> Maybe the engineers formulated and directly installed the cost-conscious principle ... in which the derived intentionality of these states would definitely lead to the human designers' own intentionality as the creators of these states. It would be much more interesting if the designers had done something deeper. It is possible ... that they designed the robot to be cost-sensitive in many ways and let it 'figure out' from its own 'experience', that it should adopt some such principle. In this case the principle would not be hard-wired but flexible ... (ibid: 71)

The more empirical response is that the pre-programming approach would never work because of the *flexibility* mentioned above – 'intelligent' behavioral skills and dispositions must be sensitive to an open-ended environment, and the only way to achieve such sensitivity is through continuous adaptation. There would therefore always be observable differences between the epigenetic and the pre-programmed robot.

On the other side of the ideological divide, someone like Searle could criticize my proposal on the basis that all the criteria it includes are *external*, and that they still leave the logical possibility that a robot which fulfills them nevertheless lacks *intrinsic* intentionality, as well as consciousness. As Harnad puts it: "there may be no one home". My response to this objection is twofold, again the first being more philosophical and the second more empirical.

First, it is not without good reasons that many thinkers, e.g. Ryle (1949), Malcolm (1971) and the earlier mentioned Wittgenstein (1953) and Dennett (1991), have opposed the idea of the logical impossibility of proving or disproving the presence of a mind in another being, known in philosophy as the "the problem of other

minds" (cf. Buford, 1971). One is that this conception of mind opens the door to a world where it is possible that all other people are zombies or robots, while all trees and stones think deep thoughts. To my knowledge it is not many people (outside lunatic asylums and philosophical departments) who hold this logical possibility to be *probable* enough worth even considering. Therefore the problem is not so much epistemological (how do I know that I am not surrounded by zombies?) but rather conceptual, showing that there is something problematic in our concept of mind which makes the problem arise in the first place. Without claiming any contribution to the philosophical discussion on 'other minds' I would nevertheless suggest that the problem arises when the concept of mind is treated as a *unitary concept with crisp boundaries*. The case could be made that instead, it is more like the concept *game* used by Wittgenstein (1953) to point out that all concepts need not be unitary arid crisp, but rather that different phenomena falling under the concept may be connected by 'family-resemblances'. Some of these phenomena involve personal experiences like *pain* but which through empathy are spontaneously applicable to other people. Other phenomena, like *shared word meaning* are primarily social, but in being grasped and followed are also individual. Since the personal and social dimensions of mind are so closely connected, the doubt in other minds is hardly more justified than the doubt in one's own.

The second response is that it is not only a matter of behavioral similarity (in the relevant aspects) that would constitute the motivation for attributing meaning to the situated embodied robot (thus making it an argument from analogy), but the possibility of a theory of how the *right combination* of situatedness, embodiment and epigenesis will give rise to human-like consciousness.[5] An example of a first attempt at such a theory is that of Edelman (1989, 1992) which views the capacity for language and consciousness as evolutionary adaptations, which at the same time require a favorable ontogenetic environment to blossom in the individual. Such a theory could show that it is impossible to have the characteristics of cognition which I proposed as criteria for the *attribution* of meaning, and at the same time to lack intentionality and consciousness: the latter emerge from the first as a matter of natural necessity. In Section 4 I will propose a developmental model which similarly implies that self-consciousness is a necessary pre-requisite for fully successful language acquisition, thus rather paradoxically offering some vindication to Turing's intuition that the *real* ability to use language implies intelligence.

3. Epigenetic Robotics and the *Cog* Project

The previous section was predominantly philosophical in addressing the *conceptual* question of the preconditions for meaning by appealing to our intuitions regarding various thought experiments on the (im)possibility of an artificial mind, and the soundness of the arguments based on them. Since I answered the question "Can a machine mean'?" in the affirmative, the task is now to suggest how an artificial, autonomous system, i.e. a robot could be possibly brought to do so.

The three characteristics of cognition used as a basis for the positive answer – situatedness, embodiment and epigenesis – would have to be fulfilled in such an implementation. On the other hand, the project of constructing such a robot can be seen as contributing to the theory suggested at the end of the last section – a theory of how intentionality and consciousness emerge in such a system – and in this way strengthen the argument for the adequacy of the conceptual answer. (It is in such loops between conceptual and empirical/constructive and back that cognitive science comes into its own.) Such a general project may be referred to as *Epigenetic Robotics.*

A more specific version of this project was recently initiated at the Department of Cognitive Science at Lund University with myself, Lars Kopp, Peter Gärdenfors and Christian Balkenius as principle collaborators. The goals of this project were formulated as follows:

- to construct a humanoid robot which is able to interact with a human "care-giver" verbally and non-verbally in a series of simple language games in a manner comparable to that of a two-year-old child
- to achieve this by endowing the robot with a (neuro)biologically plausible bodily structures, and allowing it to develop its physical, social and linguistic competence by interacting with the environment, acquiring any human language it is exposed to
- to attempt to model as closely as possible the relevant evidence from developmental psychology, i.e. certain well-established developmental progressions

Even in this somewhat simplified form, the task of Epigenetic Robotics is extremely ambitious, and many would say unrealistic given the present state of our empirical knowledge and technical resources. But if we suffer from hubris, at least we are not the only ones. A number of research labs have set themselves the task of developing humanoid robots. Perhaps the best-known one is the project centered around the development of the robot *Cog*, described by one of the researchers involved in the following terms:

> Our group has constructed an upper-torso humanoid robot, called Cog, in part to investigate how to build intelligent robotic systems by following a developmental progression of skills similar to that observed in human development. Just as a child learns social skills and conventions through interactions with its parents, our robot will learn to interact with people using natural social communication. (Scassellati, in press)

It is clear from this description that both the goals and the overall 'ideology' of this project are very similar to those of Epigenetic Robotics. A recent paper (Brooks et al., 1999) summarizes the methodological premises, achievements and challenges of the *Cog* project. The 'essences of human intelligence' according to the authors are (a) *development,* (b) *social interaction,* (c) *embodiment and physical coupling* (with the environment) and (d) *integration* (of different sensory modalities). These

characteristics clearly correspond to the three characteristics singled out in Section 2, 'integration' being an aspect of embodiment.

The features of human embodiment that have been modeled in the *Cog* project include 5 perceptual systems (visual, vestibular, auditory, tactile and kinesthetic) and a number of motor systems (2 arms, waist, neck and eyes – with altogether 21 mechanical degrees-of-freedom. There is a separate 'vision and emotive response' platform connected to a motivation system of basic 'drives' (fatigue, social contact, stimulation) which can engage in a primitive kind of social interaction, e.g. if not stimulated, its facial expression signals 'boredom', if overstimulated, it signals 'fright', otherwise it shows 'interest'. Other simple interactive behaviors that have been implemented include: *saccades* (swift eye-motion), *smooth-pursuit tracking, vestibular-ocular reflex* (stabilizing the eyes during head movement), *neck/eye orientation* ('looking' towards salient stimuli), *hand-eye co-ordination, face and eye finding*, and *imitating head nods.*

Though this list of actually implemented behaviors is impressive, one is struck by a discrepancy between the ideology of the *Cog* project and the landmarks achieved over the project's 6-year history. Notably, all of the behaviors implemented involve low-level, automatic processes, most of which are largely *innate* in human beings. Thus cognitive development is completely absent. Furthermore, since they are developed independently of each other (some on different platforms) there is hardly any *integration* between them. Finally, without the backdrop of a basic self-regulating sensorimotor intelligence, it is impossible to include any *social integration* worth its name – it is typical that the skills which are included in this category involve only crude stimulus-response based patterns.

Therefore it is not surprising when in the final part of the paper the authors identify a number of 'missing pieces'. In particular, the major problem admitted concerns the difficulty of "demonstrating coherent global behavior from the existing subsystems and sub-behaviors. If all of these systems were active at once, competition for actuators and unintended couplings through the world would result in incoherence and interference among the subsystems." (Brooks et al., 1999: 30) Other deficiencies are the absence of 'deeper visual perception' (there is nothing but feature detection in *Cog*'s visual system), and a 'sense of time' (there is nothing but very primitive procedural memory).

It is not troublesome that *Cog* does not possess these abilities – it would indeed be a miracle if it did, given the approach taken. What worries, however, is the apparent implicit assumption that somehow 'more of the same' would somehow lead to adequate coherence, recognition and memory. Such an assumption would be highly naive. A purely *reactive* system, which is what is *Cog* is at present, will not transform itself into an intentional system by just adding more input-output circuits. What is necessary is a *developmental framework* showing how such a transition could occur by extensive interaction between subsystems and between the whole system and the environment. Since despite some general pointers, Brooks and his colleagues do not offer such a framework, the task appears to be to develop one

instead. This is the foremost challenge to the Epigenetic Robotics program and in the following section I will make a first attempt to formulate such a framework.

4. Developmental Stages in Ontogenesis and Robotogenesis

The task of this section is to present the outlines of a synthetic theory of human cognitive development, and simultaneously to suggest how it might possibly be implemented in an artificial self-organizing autonomous system, i.e. a robot. In other words, I will be offering to the reader's consideration something which I may call (with a good portion of self-irony) *a recipe for the making of an artificial person.* I am quite aware that in saying so my project runs the risk of falling under the heading of neither philosophy nor cognitive science, but rather under Science Fiction. Nevertheless, there are several reasons that may be said to motivate such an undertaking.

First, as pointed out in the previous section, without a comprehensive theoretical framework, projects of building robots such as *Cog* tend to get lost – not in space but in peripheral behaviors and technical details and systematically don't live up to their lofty goals. Thus, offering a kind of "route description" for traveling from source to destination – even if turns out to be a wrong one – would serve as *a reminder of the existence of the vast territory*, and an invitation for other routes to be tried. (The only robotics project I am aware of that attempts to chart such a trajectory is that of Kozima, 1999).

Second, the *epigenesis principle* (cf. Section 3 and Section 4.1 below) stating that cognitive development progresses through a sequence of stages where each consecutive stage builds on top of the preceding one in order to yield increasing complexity, strongly constrains possible developmental scenarios. No teleportations or other mystical means for reaching the goal are allowed: The trajectory has to be *traversed*.

Third, it is possible to further constrain the developmental trajectory by relying on a multitude of ideas and evidence from different theoretical approaches and fields such as developmental psychology, social psychology, cognitive psychology, evolutionary theory and neurobiology – which nevertheless can be shown to display notable *convergences*. In particular, the work of Vygotsky (1978, 1986), Mead (1934), Piaget (1953), Gibson (1979), Bruner (1983), Tomasello, Kruger and Ratner (1993), Tomasello (1995), Neisser (1988), Changeux (1985), Edelman (1989, 1992), Müller (1995), Donald (1992) and Deacon (1997) has contributed to the developmental model presented in this section.

Finally, the *reverse engineering* method of robotics (and cognitive modeling in general) allows a considerable degree of schematicity in theorizing, abstracting away from empirical detail. Therefore, even though many of the facts and issues concerning human development remain uncertain, it is possible to proceed as if they were – and to see whether the proposed solution 'works'.

After this preamble I proceed with *Version 1.0* of the instruction manual on *How to Build an Artificial Person.*

4.1. TYPES OF INTERACTION AND DEVELOPMENTAL STAGES

The principle of epigenetic development applies to the whole life span of the individual – from conception to death. The initial state consists of the fertilized egg cell, and from then on the organism is formed as a result of a multitude of different *interactions* with the environment, with *various degree of control* from the genes. Thus, the genome does not constitute the 'blueprint' for the mature organism, but sets constraints and preferences, regulating the developmental process, which on a neurobiological level has certain similarities with evolutionary processes such as natural selection (cf. Changeux, 1985; Edelman, 1992; Müller, 1995; Deacon, 1997). From a very global perspective we can distinguish between 4 kinds of interactions responsible for giving rise to cognitive structure in the individual, with the last one applying only to human beings.

(1) *biochemical interaction*: organism-internal processes such as the "neural Darwinism" that gives rise to the basic circuitry of the brain
(2) *physical interaction*: processes through which the organism perceives and acts upon the physical environment through its sense and motor organs
(3) *social interaction*: specialized learned interaction patterns with other members of the social group
(4) *linguistic interaction*: communication by means of a conventionalized symbolic system with other members of the group and with oneself

The degree of genetic control progressively decreases from (1) to (4), and inversely, *learning* plays an increasingly important role. This much is relatively uncontroversial. What is more controversial is the following basic assumption which underlies the developmental model to be presented in this section:

> *The first few years of a child's life may be divided in 4 qualitatively different stages, each one corresponding to the period when a 'higher' form of interaction between the child and its environment assumes a dominant role for establishing cognitive structure.*

These stages are the following: From conception to birth, the human organism undergoes a *preparatory* stage in which the basic bodily and neural structures are formed through bio-chemical interactions, under close control from the genes. The stage from birth to approximately 9 months is referred to as *episodic*, attributing to this term the meaning adopted by Donald (1991), which is quite different from Tulving's (1985) notion of 'episodic memory'. The main characteristics of this stage are that the dominant form of interaction is physical and the world is perceived in terms of concrete 'episodes' (events, scenes) without any ability to conceptualize the past, the future or oneself. The next stage is called *mimetic*,

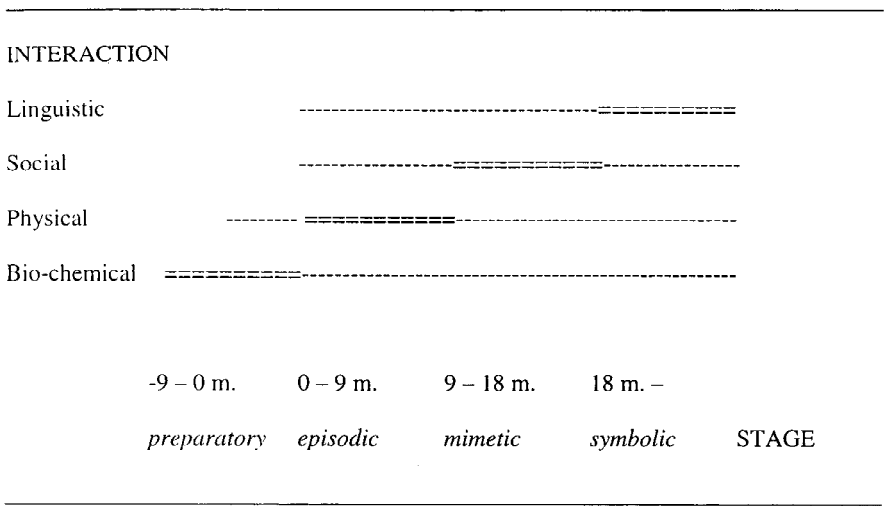

Figure 1. Developmental stages corresponding to the period (in months) during which a certain type of interaction plays the major role (marked by the double line ===) for cognitive development in the individual.

also following Donald (1991). It is dominated by social interaction – which has begun soon after birth but now truly blossoms along with the 'discovery' of others as intentional beings and the emergence of self-consciousness. Volitional acts of imitation – 'mimesis' – play a crucial role in development. The *symbolic* stage is anticipated by a 'symbolic insight', usually occurring sometime between 18 and 24 months, which marks the beginning of a new sort of memory organization, mediated by a public symbol system provided by language, as suggested by Vygotsky (1986) and Deacon (1997). The consequence of this transition are the vocabulary spurt, the emergence of grammar and a qualitatively increased role for linguistic interaction. This, in a nutshell, is the model to be presented in this section. A schematic representation is provided in Figure 1.

As reflected in Figure 1, the model assumes that the different forms of interaction proceed in parallel *all throughout development*. When a 'higher' type of interaction takes on a leading role this does not mean that the others somehow disappear. For example, even in adults, where linguistic interaction is (arguably) the most important source for the growth of knowledge, non-linguistic social interaction, physical interaction and say, hormonally driven, bio-chemical processes naturally contribute as well. The proposal is nevertheless that with each successive stage, interaction types that were dominant are superceded with a new type which is less determined by purely physical processes, and more by processes which are, broadly speaking, *intentional*. Thus, the proposed progression of stages can be interpreted as *a step-wise growth of intentionality*.

In the remaining parts of this section I will present some of the basic evidence for this division in stages, describing the cognitive structures and skills that appear

to be developed in each. Along the way, I will also consider how the most important of these could possibly be implemented in, or rather *acquired by* a robot. As suggested above, the model is partially based on the work of Donald (1991) which is actually a theory not of individual development but of evolution. Hence I will be discussing possible parallels in phylogenesis, ontogenesis, and to coin a new term – *robotogenesis*.

4.2. STAGES OF HUMAN AND ROBOTIC COGNITIVE DEVELOPMENT

4.2.1. *Preparatory Stage (−9 to 0 months)*

Empirical results strongly suggest that while still in the womb the human organism develops not only its bodily organs (along with the nervous system to support their activity) but certain specific bodily skills. The fetus has been observed to perform at least 15 different coordinated, movement patterns. Newborn babies are capable of moving their hand toward their mouth, and in the case of a miss to move both hand and head in a way that makes the hand 'reach the goal' (cf. Butterworth, 1995). While this could be (and has been, cf. ibid) interpreted as a kind of primitive intentionality (in the sense of goal-directedness), it could also be the result of a system of coordinated reflexes.[6] However, the complex co-ordinations between different sensory modalities and between sensation and motor activities that the behavior of neonates shows are indisputable and seem to require a good deal of 'innate prewiring'. Furthermore, a few weeks after birth babies display the ability to reach toward a target standing out from the background (Von Hofsten, 1989), which would require a complex co-ordination between motivation, attention, perception and motor systems. In such a co-ordination it seems possible to see the roots of intentionality: sensorimotor activity is directed to aspects of the environment which are selected by the motivation-attention complex.

The structures underlying the abilities mentioned can clearly be regarded as preparatory for physical interaction (cf. 4.2.2), but they are also indispensable for social interaction, and observable in the early appearance of imitation (Meltzoff and Moore, 1995) (cf. 4.2.3), and (later on) in feedback loops between speech perception and production, and hence for linguistic interaction (cf. 4.2.4). But there also seem to be evolutionarily established pre-adaptations in our species which are specifically 'designed' for the latter two types. These need not, and indeed for evolutionary and biological reasons most likely *could* not (cf. Müller, 1995; Deacon, 1997; Schoenemann, 1999) involve abstract structures such as an 'intention recognition module' (Fodor, 1983) or a 'language acquisition device' ("Universal Grammar") (e.g. Chomsky, 1980). Rather, likely pre-adaptations consist of much more concrete mechanisms such as the ability to detect and recognize faces (and perhaps other attributes of conspecifics) and to produce and perceive *speech* sounds (cf. Jusczyk, 1997). Finally, it is possible that a crucial pre-adaptation allowing us to cross the barrier of symbol use much easier than even our nearest relatives (bonobos and chimpanzees) is a powerfully expanded prefrontal cortex,

which according to Deacon (1997) would be instrumental to the ability to acquire symbol systems by redirecting attentional resources from immediate sensorimotor co-relations, to the relations between the symbol-tokens themselves.[7]

The need to assume pre-adaptations such as these is clearly inconsistent with a view of the new-born child as a *tabula rasa*. On the other hand, there is a gulf between this conclusion and the strong nativist picture of (almost) completely pre-specified 'mental modules', which only need to be 'triggered' by experience. A rich set of pre-adaptations designed to facilitate the child's interaction with the physical and social environment similarly imply a rejection of the view of the newborn as 'profoundly adualistic' – with no separation at all between self and environment, a view taken by e.g. Piaget (1953). Again, this does not imply a jump back to a 'natural dualism' of the type espoused by the 18th century philosopher Thomas Reid (1912), as Butterworth (1995) assumes. The self, like most, if not all, mental structures and abilities remains to be constructed in interaction, even if the child is born well-prepared for this constructive process.

What do these considerations imply for the 'preparatory stage' in the development of a robot? First, a few crucial differences need to be pointed out. While the state in which the human child is born is not simply the product of the information in the genes, but of bio-chemical and to a limited extent other kinds of interaction (cf. above) – the structures with which a robot is 'born' need to be *implanted* by its creators, i.e. they are going to be 'innate' in the strong sense. On the other hand, by experimenting with different configurations and implementations of these structures, the roboticists are taking on the role of evolution and speeding it up considerably compared to natural conditions, a point similar to that made by Dennett (1995) with respect to *Cog*.

Otherwise, apart from substituting hardware for 'wetware' it should be possible to achieve a good deal of isomorphism. An epigenetically developing artificial system should not be endowed with Chomksyan and Fodorian knowledge modules, but rather with a coordinated system consisting *minimally* of:

- organs of perception, the operation of which can easily be *coordinated*[8] to form a cross-modal network
- motor organs, which can be easily coupled with the perception network
- an attention system, initially biased for motion and perceptual contrast
- a motivation system determining the *value* of perception and motor activity in relation to internal criteria, e.g. homeostasis
- a (procedural) memory system geared to form associations between the states of the above four systems, and hence sensorimotor categories

However, while such an initial state may be sufficient for simple artificial creatures (e.g. Balkenius, 1995) it is almost certainly not enough for a humanoid robot. Given the testified capacity of neonates to form specific, short-term goals, displayed in the ability to reach out for an object and to imitate gestures, it would appear to be necessary to endow the robot with a capacity for *proto-intentionality*, i.e. a rep-

resentational ability to form goals prior to behavior, and to evaluate whether these goals are subsequently met or not. Preadaptations for *social interaction* should probably include:

- templates for human faces and/or other properties of human anatomy and motion, allowing easy discrimination of other social agents (human beings or other robots)
- extending the motivation system to include not only values directly related to survival needs, but a preference for social contact as such
- a capacity for imitation of body movements (though this could also be an emergent characteristic from cross-modal mappings, proto-intentionality and a value for social contact)

Finally, pre-adaptations for *linguistic interaction* would include, and hopefully be limited to:

- special circuitry for perceiving and producing phonetic contrasts
- the potential to be able eventually to suppress more immediate sensorimotor co-relations, and to reorganize memory on a symbolic basis (the hypothesis being that without this ability no true symbolic language would emerge, but rather something like the language skills of apes)

It should be emphasized again that none of the proposed 'innate characteristics' constitutes a complete mental module in the spirit of hard-line nativism, and neither do they imply epistemologically basic, 'Kantian' categories. All are pre-adaptations, or *preparations* for subsequent stage – hence the label for this stage of development.

4.2.2. *Episodic Stage (0–9 months)*

The period from birth to until about 9 months is a period during which the child undergoes intense development in all possible ways, to which several different kinds of interaction actively contribute. The genes regulate bio-chemical processes that lead to considerable body growth. Participation in social exchanges involving imitation (Meltzoff and Moore, 1977) and 'protoconversations' (Bateson, 1975) is observed soon after birth and develops henceforth. Linguistically, though not yet symbolically (cf. 4.2.4), the child adapts its speech perception to only those contrasts which are present in the ambient language by 6 months, and begins to experiment with speech production itself through babbling at about the same time (Locke, 1993). (These 'precocious' developments are shown by the dashed lines preceding the double lines in Figure 1.)

However, is not unreasonable to suppose that what is *most* characteristic of the infant's early development is establishing a familiarity with the (radically new, compared to the previous period) physical environment, as well as with its own body, by engaging in almost ceaseless *perceptormotor activity*. Two highly influential figures of 20th century psychology have emphasized the importance of this

kind of activity for cognition: J.J. Gibson and Jean Piaget. Though seldom brought together in this manner, I would suggest that what is probably each one's pivotal theoretical concept – *affordance* (cf. Gibson, 1979) and *(sensorimotor) scheme* (cf. Piaget, 1953), respectively – share crucial properties which makes them compatible in a theory of child (and perhaps robotic) development. In particular: they are both *ecological* notions since they can only be characterized in terms of the interaction between organism and environment – they are, properly speaking, neither 'in the mind' nor 'in the world', and they couple tightly perception and motor activity. In Gibson's terms what the infant first and foremost perceives are those aspects of the environment which *afford* some sort of action. Since the child's bodily skills change rapidly during the first year of life, so will its perception. The way objects are perceived, for example, will change once the child has become able not only to reach toward, but to touch and grasp those within reach. In Piaget's terms, the schemes coordinating perceptions and motor patterns will be the means through which objects and events are 'assimilated', i.e. perceived and categorized.[9]

The similarity between affordances and schemes may be illustrated by considering the infant's behavior towards desirable objects. As mentioned in the previous section the neonate would more or less instinctively reach toward a 'blob' in its field of vision that happens to stand out perceptually from its surroundings. This standing-out, due to strong color or hue contrast or motion, can be seen as an elementary affordance: the 'blob' affords easy detection. Next, if the child's hand manages to come in contact with this blob, i.e. it turns out to be *touchable*, the child will attempt some sort of *grasping*, probably due to a reflex. If the blob affords grasping then this will be registered, and if it furthermore affords *moving* (as a whole), and perhaps *picking-up*, then the blob would have attained many of the properties that define *objects* experientially. The next time the child sees an identical phenomenon, it will not see it as a "blob" any more – but as something graspable, movable and pick-up-able.

The schemes, on the other hand, can be seen as the sequential structures that bind the perception of the various affordances with the actions which they afford. Thus, to carry out interactions with everyday objects (such as the toys in the language game of Section 1) the infant will need:

- a gaze-focusing scheme (toward something standing-out)
- a reaching scheme (toward the same thing)
- a touching scheme (toward something touchable)
- a grasping scheme (toward something graspable)
- an object-moving scheme (toward something movable)
- an object-lifting scheme (toward something liftable)

This way of formulating the interrelationship between affordances and schemes, which to my knowledge is not orthodox,[10] has the advantage that it makes it possible to see how schemes – which rise long before concepts – can nevertheless

serve as the basis for *intentional behavior prior to the existence of explicit goals*. A developmental sequence such as the following is conceivable:

(1) In focusing attention (itself controlled by motivation) on some specific aspect of the environment, the aspect will be *seen through its affordances* (graspable, edible, etc.).
(2) Affordance-based perception can trigger appropriate goal-directed action schemes.
(3) If a simple scheme can not lead to the goal, schemes can be *chained*, e.g. the sequence of schemes leading to lifting an object listed above.
(4) Frequently chained schemes can be automated.

Since the environment is 'directly perceived' in terms of what may be called its *telic structure* (what can be done with it) initial schemes implicitly include the third necessary component of a Piagetian action scheme, namely the *expected result* (cf. von Glasersfeld, 1997). 'Accommodation' (change in the schemes) will occur if, for example, a grasping action does not lead to actual grasping since the object turns out to be too slippery. Such a 'perturbation' would require changing one's perception – the object no longer affords grasping, or the establishment of a new scheme, appropriate for slippery objects.

Furthermore, by means of sensorimotor interaction with the environment the infant not only discovers the world, but also *itself* – as a 'being-in-the-world'. This is because the child can perceive its own body, as engaged in various activities, through what Leder (1990) calls a "folding back of the body upon itself":

> Thus, my sensorimotor capabilities can reflect one upon the other; my eyes can look down upon my hands, my hands can touch each other or reach up to touch my eyes. If the body were simply a point source, an eye at the moment of sight or a fingertip at the moment of touching, it could not gain this perspective on self (ibid: 23).

The point is that each sensory modality *is blind to itself*, but when they are coordinated in loops of perception and activity, an *awareness* of one's body, as separate from the world, emerges. The term proposed by Neisser (1988, 1995) to denote this initial self-awareness is *ecological self*: "the individual considered as an active agent in the immediate environment. Such agents perceive themselves, among other things: where they are, how they are moving, what they are doing, and what they might do, whether a given action is their own or not" (ibid: 4).

There is a close similarity between this notion of a self which acts here-and-now (the modal "might" in the quotation above is questionable) but remains unreflected upon and Edelman's (1989, 1992) notion of *primary consciousness*. This is the kind of awareness of the world that an individual with only an ecological self would possess: "the state of being mentally aware of things in the world – of having mental images in the present. But it is not accompanied with any sense of a person with a past and a future" (ibid: 112). According to Edelman, most higher animals

– and clearly primates – would possess such primary consciousness, the crucial evolutionary advantage of which is to provide the organism with the ability to see the world not in terms of disconnected stimuli or even objects, but as coherent *scenes*:

> Learning certainly occurs in animals that show no evidence of conscious behavior. But in some animal species with cortical systems, the categorizations of separate causally unconnected parts of the world can be correlated and bound by a *scene*. By a scene I mean a spatiotemporally ordered set of categorizations of familiar and nonfamiliar events, *some with and some without necessary physical or causal connections to other in the same scene*. The advantage provided by the ability to construct a scene is that events that may have had significance to an animal's past learning can be related to new events, however casually unconnected these events are in the outside world (ibid: 118).[11]

Despite using some of the same terms differently, Donald (1991) appears to be aiming at the same (or at least a very similar) notion, by proposing that what characterizes "ape culture" in contradistinction to human is that it is entirely *episodic*:

> Their lives are lived entirely in the present, as a series of concrete episodes, and the highest element in their system of memory representation seems to be at the level of event representation. (ibid: 149) . . . Event perception is, broadly speaking, the ability to perceive complex, usually moving, clusters and patterns of stimuli as a unit. (ibid: 153) From a human viewpoint, the limitations of episodic culture are in the realm of representation. Animals excel at the situational analysis and recall but can not re-present a situation and reflect on it, either individually, or collectively (ibid: 160).

This close analogy between the main type of representation used by apes, according to Donald (1991), and that employed by children during the first 9 months of life, is what motivates using the rather unusual label 'episodic stage' for this period of human development. This general similarity between stages in evolution and in ontogenesis follows not from an unmotivated principle of 'recapitulation' (Haeckel, 1874), but from the principle of epigenesis itself. It implies that there should be an analogy between phylo- and ontogenesis for the following important reason: 'Simple' forms of representation and their supporting neural structures served undeniably as the 'stepping stones' on which evolution built more complex representations and structures which otherwise would have been impossible. While *Homo Sapiens* may be pre-adapted for such more complex representations and language, there is no reason to think that we have lost the capacity for the 'simpler' intermediary ones. And since these served a supporting function in evolution, there is every reason to think that they play a similar role for individual development. The fact that these simpler forms resurface as 'cognitive vestiges' when the more complex ones give way (cf. Donald, 1991) supports this view.

How would it be possible for a robot to develop an 'episodic mind'? *Cog* is obviously very far away from achieving any of the competencies discussed in this section, but *Cog* also lacks the structures of the preparatory stage, notably the coordinated system coupling perception, action, attention, motivation and memory. If a robot with such a system, suitably 'pre-tuned', were exposed to an environment it can interact with, it would constitute an autonomous system in which 'bodily skills' of increasing complexity may begin to develop. In particular, we may expect it to be able to:

- achieve stable coordination between the direction of its gaze and reaching
- distinguish objects from their background
- pick up affordances and through them recognize an increasing number of aspects of the environment
- learn sensorimotor schemes related to these affordances (e.g. look in different directions, move in different directions, reach out and grasp, lift up and drop)

Of course, many questions remain on how to construct an artificial system capable of such relatively "high-level" perception, and how to combine it with the ability to act on these perceptions, making accommodations in the case of failure. An interesting proposal for how to implement a 'schema mechanism' has been suggested by Drescher (1991) while Balkenius (forthcoming) offers ideas for a biologically realistic model of attention. The crucial problem for an autonomous robot, however, seems to be that of motivation. Without it, there is nothing to make any aspect of the environment more desirable than any other, to focus attention, and to recognize a successful assimilation from a perturbation, i.e. when the (implicitly) expected result obtains, and when it does not.

It is not yet known how the human motivational system works, but it is clear that 'primary conditioners' such as hunger, thirst, pain, and pleasure play an important role. If these are crucial for the development of human (and animal) intentionality, then an artificial system will be, by definition, inadequate. But it is possible that motivations less directly based on survival play a central role in guiding human behavior and these could be more amenable to robotic implementation. Similar to Hebb's (1955) proposal of an "optimal level of arousal" a robot could perhaps be designed to strive for an 'optimal level of novelty'. Such a homeostatic-like "generalized drive" could regulate behavior in the following way: too low degrees of stimulation would lead to a motivation state of 'boredom' which stimulates active search for novelty; too high degree of novelty leads to 'fear', causing inaction, or a search for a situation with less novelty; while an in-between state would lead to a state of 'interest' and a sustained exploration of the environment similar to that of tranquil infants. Guided by its generalized optimal-novelty drive such a robot could posit (for itself) short-term goals, such as *explore-that-new-object* and accomplish them by a succession of schemes. Furthermore, it should be able to learn to distinguish between itself and the environment and have a kind of 'body image' that would be a correlate to an ecological self. One consequence would be

that it knows the limits of its 'reaching space'. All these are skills that can be quite unambiguously tested behaviorally.

What remains more mysterious is whether such a robot could be expected to possess any form of "primary consciousness". If we assume Edelman's characterization to be more or less correct, we could express the following tentative hypothesis: Primary consciousness could arise in an autonomous robot of the type suggested in this section if (a) it is advantageous for it to have representations of whole scenes (episodes) and (b) the world is perceived through such representations via "reentrant mappings". In the sense in which primary consciousness, as well as qualia, are first-person phenomena it remains impossible to verify their presence objectively – in an artificial and natural system alike. However, if a robot could display most of the characteristics of a being with an 'episodic mind', after having achieved these through a developmental process which resembles that of neonates, there seems to be no principled reason to exclude the possibility of their existence.

4.2.3. *Mimetic Stage (9–18 months)*

As pointed out earlier, social interaction does not wait for the child to become familiar with her physical surroundings and bodily self. Neisser (1988, 1995), for example, even proposes that an early, directly perceived *interpersonal self* develops alongside the ecological one.[12] However, despite the obvious existence of social skills in infants prior to the age of 9 months, documented in the research of Meltzoff, Trevarthen and others, there is a clear change in the young child around that period, characterized by Tomasello (1995) in the following dramatic terms:

> At about 9 months of age, infants begin to behave in a number of ways that demonstrate their growing awareness of how other persons work as psychological beings. They look where adults are looking (joint attention), they look to see how adults are feeling toward a novel person or object (social referencing), and they do what adults are doing with a novel object (imitation learning). . . . Infants also at this time first direct intentional communicative gestures to adults, indicating an expectation that adults are causal agents who can make things happen. All of these behaviors indicate a kind of social-cognitive revolution: At 9 months of age infants begin to understand that other people perceive the world and have intentions and feelings towards it; they begin to understand them as intentional agents (ibid: 175).

The crucial feature of Tomasello's proposed 'social-cognitive revolution' is therefore an *explicit awareness* (consciousness) of first others, and then of oneself, as intentional agents. This awareness is qualitatively different from the ecological self and primary intersubjectivity of the first period. It is not yet a complete "theory of mind", considering that it is not until about the age of 4 that children begin to ascribe right or wrong beliefs to other people (Harris, 1991). Still, it is clearly a

kind of "intentional stance" that the children take, not in a pick-and-choose-manner as Dennett (1987) would have it, but as *an inevitable consequence of striving to make their social world comprehensible*. By interacting with these "psychological beings" on a par, first communicating through "significant gestures" prior to the acquisition of language the child develops a 2nd-person perspective of itself, a "me" (Mead, 1934; Vygotsky, 1978) which implies (proto) self-consciousness. Possessing such self-consciousness offers a new type of representation to cognition, which can serve in predicting others' behavior, planning one's own and comparing one's behavior with that of others much more sensitively than before. Since the behavioral evidence strongly suggests that the child achieves such self-consciousness prior to the acquisition of language, this representation can not be predominantly linguistic. Even more – it seems highly likely that without it language acquisition, which requires a sensitivity to shared social norms, would be impossible.[13] Following Donald's (1991) proposal for a similar stage in hominid development, this period in ontogeny and the representation it brings may be called *mimetic*.

Again, there are substantial parallels between phylogeny and ontogeny which motivate the present use of the term for child development. According to Donald, a diverse corpus of evidence – the signs of the relatively complex culture of (apparently) non-linguistic *Homo Erectus*, "self-contained" mimetic vestiges in certain neurological patients, and the high cognitive capabilities of (yet) non-linguistic individuals, such as pre-verbal children – suggest the existence of an "archaic but distinctly human culture that mediated the transition from ape to human. This mediating, or intervening, layer of hominid culture is called mimetic, on the basis of its dominating, or governing mode of representation" (ibid: 162). Donald lists a number of properties of mimetic representation: *intentionality* (e.g. the ability to attribute intention to the gaze of another, seemingly missing in chimpanzees), *generativity* ("parsing" motor acts and recombining them creatively), *communicativity* ("mimetic acts by their nature are usually public and inherently possess the ability to communicate" ibid: 172), *reference* (separation of represented from representation) and *autocuing* (voluntary recall and execution). To be able to perform such acts of mimesis, attention has to be carefully controlled so that it may shift from one's own body as a signifier (representation), and whatever is being signified (represented). This implies a reflective, objectified relationship to one's body, and the step from this to self-awareness – at least as a sensorimotor being – is small. Thus, the ability to engage in acts of mimetic representation itself appears to imply self-consciousness.

However, while Donald focuses on the kinds of advantages mimesis offers once in place, he does not take on the burden to suggest a scenario of *how mimetic skill emerged*. This leaves an obvious gap in the argumentation, since it is not clear where the 'self-representations' required for mimesis would derive from, given that the 'episodic mind' of primates is not obviously self-conscious.

There are indications that at least some primates do possess some form of a self-concept (a 'me'): in Gallup's (1970) well-known *mirror test* experiments chimpanzees learned to recognize themselves after a while, orangutans more slowly and gorillas not at all. Recognizing oneself in a mirror may be said to involve being able to match the ecological self (the body as felt and seen here and now) with a 2nd-person view of oneself, and this appears impossible to be grasped as such without a self-concept.[14] Interestingly, these species are not only the closest to us genetically, but also socially – their groups possess complex hierarchical structures maintained through e.g. grooming (Dunbar, 1996). Furthermore, if a chimpanzee has been reared in isolation, it fails to pass the mirror test as an adult (Gómez, 1994). All this suggests that Donald's evolutionary scenario may be bolstered by an argument similar to the one presented for ontogenesis: in brief, the increased social complexity of the culture of *Homo Erectus*, required e.g. for the successful spread throughout Euro-Asia, lead to increased importance of mechanisms for social co-ordination such as joint attention and pointing, and consequently to an enhanced self-concept. Gómez (1994) offers an intuitive scenario for how these could be related, applicable to our ancestors as well as to our children:

> When the attention of another person is focused on you – that is on your body – you can get oriented to yourself as a physical entity. However, when the focus of another person's attention is your attention, this may act as a clue toward your own attentional activity. The other's attention points to your own attention and, as a result, you are led to your own attention as focus of attention. If we consider the signs of attention as signs of awareness, the structure of attention contact seems to lead to a first version of self-awareness, both as a physical and as an "aware" or "attending" entity (ibid: 76).

Thus, requiring at least a rudimentary form of self-consciousness, mimesis was most likely itself a product of increased social interaction (though tool-use can not be excluded as a contributing factor). When established – through a combination of learning and natural selection – it could exercise further consequences for social organization by allowing feats such as complex imitation, gestural communication, ritual and pedagogy, thus giving rise to the "mimetic culture" hypothesized by Donald.

As far as children are concerned, many of the skills that characterize infant cognition during this stage such as imitation, pointing and joint attention (as well as the first uses of language) may be seen as essentially involving mimesis. The advantage of positing a mimetic stage in development, during which non-linguistic social interaction plays the leading role, is that we can both make better sense of the impressive cognitive achievements of pre-verbal children – without resorting to Fodorian nativism – and are in a position to provide a better account of the onset of (symbolic) language in the next stage. As in phylogeny, the appeal of a mimetic stage in ontogeny is that it serves as a 'missing link'.

Considering that even the competencies of the episodic stage described in Section 2.2.2 are far beyond the scope of present day robotics, the ones described in this section are bound to strike practicing roboticists as hopelessly difficult. Perhaps so – but this makes them no less worthy of their consideration – if they indeed intend to model human intelligence and language. Those not aware of these 'difficult aspects' are bound to fall into the kind of short-sightedness characteristic of the *Cog* project and e.g. attempt to model language acquisition in terms of associations to perceptual stimuli and motor programs – the kind of cognitive skills that exhaust the competence of an insect. Since language presupposes much higher cognitive capacities (including self-consciousness) present-day efforts to 'implement language' in robots are about as likely to succeed as a project aiming to teach language to an ant!

Therefore, despite that it is at present technically inconceivable how a robot could be brought to the degree of self-consciousness to allow mimesis and vice versa (Zlatev, 2000) – it appears to be the only way to acquire true meaningful language. A possible way to achieve this goal in an artificial system is to provide some kind of equivalents to the forms of social interaction characteristic of children, attempting to lead the robot through a similar developmental path. Three of the cognitive skills of young children mentioned above – imitation, pointing and joint attention – undergo step-wise development and their development could perhaps be subjected to modeling within an epigenetic robotics framework.

The modeling of *imitation* would constitute a good starting point for any attempt to build 'artificial persons'. On the one hand, it is clear that it has an innate component, testified by the abilities of neonates to imitate behaviors such as tongue protrusion, mouth opening, blinking and emotional expressions (Butterworth, 1995) and possibly related to "mirror neuron" systems for cross-modal mapping (Rizzolatti and Abrib, 1998). However, even during the episodic stage imitation is not just a matter of mimicry, a literal copying of a behavior in the manner of parrots. Meltzoff and Moore (1995) list four features of infant imitation which distinguish it from mimicry (and possibly even from the imitation of other species): *intentionality* (ability to distinguish between the goal state and the actual performance and correct the latter in terns of the first), *selectivity* (imitating different aspects of the model), *creativity* (original ways of achieving the same goal), *volitional control* (infants may defer imitation for a later moment or rather chose not to). While it is somewhat debatable whether the imitations of babies prior to 9 months possess all these features, they clearly do so during the mimetic stage, when one's own imitative act can become the focus of explicit attention. Furthermore, it is possible that a more rudimentary type of imitation may itself be the key to the rise of concepts of self and other, as suggested by e.g. Baldwin (1913: 134): "My sense of myself grows by imitation of you and my sense of yourself grows in terms of myself." The idea seems to be that imitation functions as a kind of bridge of intersubjectivity between self and other, along which properties of the other can be transferred back to "me" – and thus help create the self-concept – and properties

of the ecological self such as pain are transferred to the other, and allow for the emergence and development of empathy. The conclusion would be that if for one reason or another imitation is impaired, this should lead to difficulties in both these processes – which is precisely what is observed in *autism* (Frith, 1989; Kozima, 1999). Tomasello (1995) furthermore conjectures that different degrees of autism (and Asperger's syndrom) could be correlated with the (developmental) age of the child when the insult (genetic or acquired) occurs. If a robotic model of imitation could be realized, hypotheses such as these could be empirically tested.

Pointing is a form of gestural communication which develops during the mimetic stage, and while it may be partially based on imitation (as assumed by the *Cog* researchers), it could alternatively (or also) arise from the child's non-communicative attempts of reaching a desired object in combination with the caregiver's attribution of communicative intent, as suggested by e.g. Vygotsky (1978). Around 9 months the child's reaching becomes stylized and is accompanied by a gaze to the adult's face. By 12 months the child is practicing so called *declarative pointing*: the arm and index finger are clearly extended and often pointing is done with the intention of drawing the adult's attention for its own sake, e.g. pointing at the airplane or at the moon. Some of this developmental structure could perhaps be implemented in a robot, motivated by e.g. novelty as proposed in Section 4.2.1. It could 'learn' that desired novel objects can be obtained just as efficiently by pointing at them (and perhaps vocalizing) when there is a caregiver in the vicinity. But it is unclear what would ever bring it to engage in declarative pointing, and if so its pointing behavior would remain (almost) purely instrumental, similar to that of certain trained apes, including the famous Kanzi (Savage-Rumbaugh et al., 1998).

Joint attention is a reflexive process in which two individuals are not only attending to the same thing, but are also aware that they are attending to the same thing. This obviously implies at least a degree of awareness of self and other as "psychological beings" (cf. Tomasello's citation above). Like pointing, however, the full capacity for joint attention seems to develop gradually, leaving the possibility for this developmental structure to be more or less successfully implemented in a robot. The first step is the establishment of eye-contact, which the infant accomplishes by the age of 2–3 months. By 6 months the child begins to follow the general direction of the other's gaze and around 9 months it can generally trace the adults gaze to the *first* salient object, by 12 months to other objects and by 18 months to objects outside the child's own field of vision (cf. Butterworth, 1991).

Even from this very superficial description, it can be seen that there is a very close connection between joint attention and imitation. In fact, joint attention may even be considered an instance of imitation, once the intersubjective nature of imitation is appreciated. The similarity persists in the developmental structure as well – from more literal and rigid to more intentional and context-sensitive. At its highest stages, it indicates an assumption of a common world of experience, aspects of which can enter a window of intersubjective attention – which appears

to be a necessary prerequisite for the development of language (e.g. Bruner, 1983), a theme continued in the next section.

Finally, assuming that the hypothetical robot has managed to undergo the kind development in its acts of 'social interaction' described in this section, showing child-like skills of imitation, gestural communication and joint attention, even though as yet little or no language, it would make sense to expose it to the Mirror Test. If it passes it, the hypothetical team of roboticists should probably start considering ethical questions regarding future experimentation.

4.2.4. *Symbolic Stage (18 months –)*

After discussing how different pre-linguistic cognitive structures can arise from bio-chemical, physical and social interactions, it is finally time to turn to the kind of interaction which is centrally implicated in the use of language – linguistic *cum* symbolic interaction. The central questions to address are: How do children acquire *a system of conventional (shared), mostly arbitrary signs*, which is one of the traditional, and largely adequate definitions of language (e.g. Saussure, 1916)? And furthermore, can this process be not just *simulated* (anything can be simulated), but more or less equivalently *realized* by an artificial (robotic) system so that this system will be in a position to *mean* what it says and to *understand* what is said to it?

The answers to these questions have been considerably delimited by the theoretical assumptions and conclusions making up the developmental model presented thus far. I will try to show in the present sub-section that these assumptions and conclusions – such as the development of imitation and self-consciousness during the mimetic stage – are if not necessary, then at least quite plausible. A major motivation behind them is that they offer a more promising approach to the enigma of first language acquisition than the numerous alternatives and could thus be inferred through an inference to the best explanation.[15]

One of the many empirical puzzles involved in child language acquisition is the following: On the one hand, language acquisition may be properly said to begin even prior to birth (in the preparatory stage), since children are born with the ability to differentiate not only their mother's voice from other voices, but also their "mother tongue" from other languages. On the other hand, it is not until around 18 to 24 months that most children begin to communicate with adults and peers through what is clearly recognizable as language. It is then that a phenomenon known as the *vocabulary spurt* occurs (the child's vocabulary begins to increase rapidly) and somewhat later the child's utterances increase in length to include several words, often in novel combinations. Why this 'delay' in the onset of language production? The child's babbling from 6–8 months and its first apparently meaningful expressions at about 9 months or so (Bates, 1979) show enough phonological variation to indicate that the problem is *not* in the lack of pronunciation skills (cf. Jusczyk, 1997). Furthermore, children's speech perception is quite developed by the age of 7–8 months and as we observed in the previous sections they are

extremely good imitators (once they have the relevant body organs under voluntary control). Finally, children seem to *understand* much of the language addressed to them at least from 12 months onward, but in any case prior to 18 months. So – why does not language production begin earlier and why, when it comes along (after the limited number of expressions used before the spurt), does it make such a dramatic entrance?

A more conceptual puzzle is the following: Communication (conceptually) requires *shared meaning*, not only "the same kind", but "the same instance" (cf. Itkonen, 1983). For Speaker A and Speaker B to share a particular conventional meaning X, they must have made it *intersubjective*, i.e. that they have made it an object of *common knowledge*. This implies *minimally* that A not only knows X, but knows that B knows X and similarly: B knows X and knows that A knows X. For more complex social interaction – third, fourth and higher level common knowledge may be necessary (cf. Clark and Marshall, 1981; Itkonen, 1997). What is puzzling is the question: How does the child acquire such common (linguistic) knowledge, in general, and especially in early development, when its ability to hold explicit higher-order beliefs is still limited?

The present model suggests that both of these puzzles may have a related answer: the vocabulary spurt marks the onset of *the symbolic stage* in development, which is characterized by a reorganization of semiotic activity from indices and icons to symbols, which are both more conventional and systematic. To explicate, during the first 9 months of life (the episodic stage) there is no real basis for semiotic activity and hence for communication proper, while this does not prevent certain forms of social interaction such as protoconversations and turn-taking from occurring. Certain clearly non-representational aspects of language such as greetings (*Hello-Hello*) may have their basis in essentially episodic, non-semiotic social competence. With the mimetic stage the situation changes: In differentiating self from other, and both from the world, towards which self and other stand as intentional, psychological beings, it becomes meaningful to communicate, i.e. to use signs. Using Peirce's (1931–35) three-part division of signs, children's first communicative expressions between 9 and 13 months can be seen as *indexes* (based on contiguity) – for example, pointing gestures and simple deictic expressions *here, there, look, come*. Somewhat later come what may be regarded as *icons*, not only in the sense that the signifier resembles the signified (*bow-wow*) which are quite limited in number and role, but in the sense that the words are associated with perceptual templates ("image-schemas") which resemble their referents. In this broader sense, the application of the word *ball* to all kinds of round things, or *daddy* to all adult men may be regardled as iconic. What is typical for these early word meanings is that

(1) they differ from the adults' (conventional) meaning through under-extension and over-extension and are hence not shared,
(2) the relationship between sign and meaning is not arbitrary – mimesis itself is iconic *par excellence*, and

(3) they do not participate in a system in which they are *related to each other through internal relations*.

In contrast, when the child enters the symbolic stage, somewhere between 18 and 24 months, his meanings become considerably more

(1′) conventionalized,
(2′) based on an arbitrary relationship between signifier and signified, and
(3′) connected within a system based on *internal relations*, i.e. not via the relations that hold between the things they stand for.

Point (3′) is usually not considered in discussing the differences between pre-symbolic and symbolic meaning, but it has been strongly argued for by Deacon (1997) on the basis of evolutionary theory, primate research and neurobiology. The major difference between pre-symbolic (which Deacon rather incorrectly lumps under the term "indexical") and symbolic meanings according to Deacon is that the latter "form a closed logical group of logical possibilities. Every combination and exclusion relationship is unambiguously and categorically determined" (ibid: 86) and "[i]ndividual indices can stand on their own in isolation, but symbols must be part of a closed group of transformations that links them in order to refer, otherwise they revert to indices" (ibid: 87). While primates can be brought to discover such systematic relationships between symbol-tokens only after rigorous special training (with the possible exception of Kanzi who was immature enough to pick up some of it spontaneously, cf. endnote 5), children do so naturally, relying, according to Deacon on the enhanced ability to focus attention on higher-level co-relations, deriving from a strongly expanded prefrontal cortex. Doing so allows children to build a new sort of semiotic system, which once internalized also functions as a new sort of memory system with far-reaching consequences. The transition from the indexical-iconic to the symbolic form of organizing semiosis and memory is drastic enough to resemble an 'insight':

> Although the prior associations that will eventually be recorded into a symbolic system may take considerable time and effort to learn, the symbolic recoding of these relationships is not learned in the same way; it must instead be *discovered* or perceived, in some sense by reflecting on what is already known. . . . What we might call a *symbolic insight* takes place the moment we let go of one associative strategy and grab hold of another higher-order one to guide our memory searches (Deacon 1997: 93).

An effect of this new strategy is the ability to define words in relation to each other, which according to the present model is the major cause of the vocabulary spurt, as well as the grammar spurt, which follows. Furthermore, when the tokens are interrelated, they provide strong constraints on each other's extension – if the child tries to figure out the meaning of spatial morphemes such as *in, under, through* on their own, it is bound to come up with huge overextensions, but if it can learn

them as part of a system of paradigmatic and syntagmatic relationships (e.g. in combination with the verbs they co-occur with) that would considerably help her to zero in on the conventional usage – as demonstrated in certain experiments with connectionist modeling (Regier, 1996; Zlatev, 1997). What I am suggesting is that conventionalization does not literally imply the existence of explicit mutual knowledge ("I know that he knows that I know . . . ") but an *intentional calibration* of one's own usage to match that of others, constraining the outcome of this process by symbol-internal relationships.

Why does the transition into the symbolic stage occur at this particular age of the child? While a nativist will opt for a maturation-based explanation, this is by no means necessary. There are a number of other factors that should be considered as possible causes first. In particular: the accumulation of non-symbolic expressions, both single words and whole formulaic utterances (which from the standpoint of the child are the same) would pose a pressure on memory to find some more efficient system of organization. Locke (1993) proposes a similar scenario, except that he believes that the memory pressure triggers the innate "generative device".

A second possible reason is more directly connected to the development of the child's understanding of intentionality, which is a prerequisite for the ability to use *conventions*, linguistic or otherwise. Even if the inter-symbolic relationships constrain the child's choices, and the child need not make explicit higher-order beliefs about common knowledge, it must nevertheless be able to 'home in' on the adult's *intended* meanings, which are not obvious from the physical setting alone. And it is characteristic that only around 18 months the joint attention complex seems to be firmly established (cf. Section 4.2.3) signaling that the child has adopted an "intentional stance" toward others and self (most probably in this order), as e.g. claimed by Meltzoff and Moore (1995):

> Our research suggests that by 18 months of age, infants are not strict behaviorists. They ascribe goals to human acts. Indeed, they infer the goal of a sequence of behaviors even when the goal was not attained. They do this in preference to literally reenacting the motions seen. Thus it appears that they code the behaviors of people in psychological terms, not purely as physical motions (ibid: 61).

As argued in Section 4.2.3 this type of enhanced intentionality, one of the ultimate achievements of the prior mimetic stage, is impossible to imagine without at least a limited form of self-consciousness. Hence, reflective consciousness can not really be a *consequence* of, but must rather be the *pre-reguisite* for proper, meaningful, language use – in opposition to theories of consciousness which claim that language is necessary for any kind of (self-)consciousness to appear, such as those of Edelman (1992) and especially Dennett (1991).

On the other hand, due to the mediational role of signs for thinking (e.g. Vygotsky, 1978) the vocabulary spurt also implies a conceptual spurt. Symbolic memory offers "the ability to elaborate, refine, connect, create, and remember great numbers

of new concepts." (Edelman, 1992: 130). It is therefore quite probable that with the onset of (symbolic) language acquisition a rather new type of consciousness emerges. One characteristic of the 'symbolic mind' would be a strongly *empowered imagination* – the ability to create realities different from ongoing experience. Even if this is not the first release from "the tyranny of the remembered present", as Edelman would have it, it would considerably facilitate, and maybe even make possible for the first time *explicit* concepts of past and future. Even primary ('episodic') consciousness must have a temporal dimension – since any kind of phenomenal experience is unthinkable without it – but it hardly stretches to the distant past, and even less likely to the yet unforeseeable future, which must be products of the imagination. Similarly, mimesis could serve the basis for acting out a 'story' of some complexity (especially given conventions for interpretation), but it is difficult to see how one could 'mime' creation myths (Donald, 1991) or life stories which are probably constitutive of autobiographic memory (Bruner, 1986; Neisser, 1988). Both are forms of *narrative*, which are made possible by the complex structure of paradigmatic (hierarchical) and syntagmatic (sequential) relationships available in language. Finally, language makes possible the ubiquitous phenomenon of *internal speech*, which has been argued to be functional for problem solving (e.g. Vygotsky, 1986) and which Carruthers (1996, 1998) argues is constitutive for conscious, propositional thought, i.e. implying that thought processes that are independent of language are either unconscious or non-propositional or both.

In sum, once linguistic interaction becomes symbolic, it assumes a central role for the cognitive development of the child and a host of new cognitive abilities follow. These include a strongly empowered imagination, concepts for past and future (and counterfactual scenarios), narrative skills and internal speech. Even though it seems unreasonable to claim that these give rise to either conscious experience or reflexive self-consciousness *per se* (*pace* Dennett and Edelman), they constitute a conceptual revolution which makes us unique in the natural world, for all that we know. Is there ground to think that we could lose this uniqueness, as a result of development in the artificial world?

Assuming that the hypothetical robot this paper has been constructing 'virtually' has, despite all difficulties, been able to achieve the cognitive landmarks of the mimetic stage – imitation, joint attention, intentionality and (even) self-consciousness – it is not unreasonable to answer the above question in the positive. Conversely, the reason for the persistent failures of all "natural language processing" and "language acquisition" artificial systems *aiming at human-like competence* (i.e. excluding some very successful systems which simply 'process' language for particular practical purposes) is easy to explain – they have all started from the wrong end (cf. Dreyfus, 1993; Kozima, 1999; Gärdenfors, 2000). Symbolic linguistic interaction builds on non-symbolic forms of interaction and cognition ('embodiment'), requires grounding in sociocultural practices and conventions ('situatedness') and, as I have repeatedly suggested in this paper, at least a simple form of self-consciousness. None of these properties, and even less so

their combination has been developed to any significant degree in any artificial intelligence system up to date, 'symbolic', 'connectionist', 'hybrid', 'robotic' or whatever other, and therefore all computational models of language have been no more that *toy models*.

But it would also be wrong to assume that if one had managed (somehow) to install mimetic competence in a robot, language would just fall out automatically. The huge tasks of language acquisition and meaningful language use would still remain, only this time they may be approached realistically, that is, not *too* far removed from the starting point of the child. Thus, language should be able to emerge in the robot under conditions similar to those under which it emerges in children.

What more precisely is one to expect from the epigenetic robot linguistically? In the same way that it was possible to formulate behavioral tests for evaluating the robot's development with respect to physical and social interaction, we can appeal to standard criteria for behavioral success in language acquisition, comprehension and production, when judging the robot's linguistic-symbolic competence. With respect to *comprehension* we may evaluate the robot's ability to deal with language on the basis of criteria similar to those which Leneberg (1980) has formulated for accrediting the language skills of apes as qualitatively similar to those of *Homo Sapiens*.

- linguistic interaction with the robot should be carried out through whole utterances, rather than just single words
- this interaction should result in the ability to recognize at least a limited vocabulary (e.g. 100 items)
- questions and requests consisting of any grammatical and meaningful combination of items from the limited vocabulary should be comprehensible
- an ability to respond to utterances including novel words, and to learn the meaning of these new words quickly

The last criterion obviously involves *language acquisition*, which can not be separated from language use in general. More specific criteria with respect to the structure of the acquisition process may include the following:

- comprehension should in general precede production – there are many contextual cues which may be utilized for the sake of the first, while the latter requires more specific symbolic and grammatical knowledge
- a "pre-symbolic stage" with many under- and overextensions should be observed prior to the vocabulary spurt
- the vocabulary spurt may involve either more nouns or verbs depending on their proportion within the speech of the "caregiver" (for differences between English and Korean children, cf. Gopnik, Choi and Baumberger, 1996)
- the vocabulary spurt should be soon followed by "grammar takeoff', a parallel increase in the complexity of productive utterances (Bates, in press)

- there should be easy recovery from mistakes in word-meaning and grammar, and a sensitivity to the norms of proper usage (without the need for explicit correction)

Finally, with respect to *language production* the following set may be viewed as a developmental ladder leading into such heights of cognitive complexity that if reachable, the robot's status as an artificial person should be quite clear.

- an ability to use language instrumentally, e.g. to form requests for the fulfillment of desired states-of-affairs, or to ask about the whereabouts of desired objects
- an ability to use language epistemologically, e.g. to use declarative utterances to communicate states-of-affairs and to test hypotheses concerning one's own knowledge
- an ability to engage in self-directed speech, e.g. accompanying problem solving
- an ability to engage in self-reflective speech, a form of self-conscious thinking, e.g. *cogito ergo sum*

Finally, it is instructive to contrast these criteria for evaluating the 'intelligence' of an artificial system with the Turing Test, cf. Section 2. First of all, they are not meant at all as any kind of 'operational definitions' of linguistic competence, but only as behavioral evidence for the possession of such competence. Second, they are only the tip of the cognitive iceberg, assumed to rest on a mountain of bodily and social skills without which they would be unthinkable. And third, they are to be achieved through a naturalistic developmental process, while the TT placed no such constraints allowing the 'intelligent' system to have been simply cleverly programmed. The combination of these differences allow us to treat these criteria as evidence for the presence of genuine, intrinsic intentionality and meaning in a machine – while the TT does not.

4.3. SUMMARY

In sum, the developmental model presented in this rather long section was built on available ontogenetic evidence, phylogenetic evidence, and a sizable portion of abductive reasoning (i.e. speculation). The red thread connecting the different pieces was the principle of epigenesis stating that during every state of development, new structure arises on the basis of existing structure plus various sorts of interaction. Physical, social and linguistic interaction intertwine since birth in playing such a role, but I suggested that they have their peaks in three consecutive stages of development: episodic, mimetic and symbolic. The transitions between these stages are qualitative, and bare a similarity to the stages in phylogenesis, proposed by Donald (1991) and Deacon (1997). The epigenesis principle makes it reasonable to find such similarities between ontogeny and phylogeny, since preadaptations for higher cognitive functions rely more or less directly upon evolutionarily simpler structures

and processes, and while the first are in the process of being (re)-constructed, the latter dominate the early stages of ontogenesis.

While it may seem utopian to expect that an autonomous artificial system – a robot – could undergo the same, or similar development, I have suggested that there does not seem to be any reason why it could not, in principle. It would of course have to be adaptable enough, and have the potential for the types of physical, social and linguistic interaction that the child has, and this raises enormous technical difficulties, but as far as I can see, no unsurmountable ones. Following the principle of epigenetic development, robotogenesis could *possibly* recapitulate ontogenesis, leading to the emergence of intentionality, consciousness and linguistic meaning.

5. Conclusions

In this paper I have presented a rather long, and somewhat loose, argument for a particular type of answer to one of the central questions of cognitive science – *Can a machine have meaning and if so, how can this be achieved?* – which I will now try to summarize.

The first part of the question raises primarily conceptual issues and thus requires a, broadly speaking, *hermeneutic* treatment. The second part is in itself predominantly *constructive*, i.e. technical. But since the most reasonable approach to solving it is through 'reverse engineering', i.e. by analyzing a specimen who has the required properties and asking how a replica of it 'in all the essential aspects' can be constructed – the question becomes *empirical* too: how does that specimen actually function? Each one of these separate questions is so huge that one is at a loss where to begin. To consider them *all together* may seem hopeless. Nevertheless, there is a good reason to do so – *they constrain each other*. The hermeneutic question logically precedes the constructive question: we need to know *what* it is we want to achieve, before asking *how* to achieve it. On the other hand, a successful demonstration of a possible answer to the constructive question would offer an existence proof for (at least one interpretation of) the hermeneutic question. An answer to the empirical question, as mentioned, would provide a blueprint to be followed in solving the constructive question. On the other hand, in constructing a model, certain empirical predictions can be tested and others for the first time discovered. In sum, the allure of cognitive science lies in the assumption that by bringing together conceptual (hermeneutic), constructive (modeling) and empirical (scientific) questions and methodologies on the enigma of the nature of the human mind, an answer to the whole complex would be more revealing than an answer to its parts. This assumption has motivated the seemingly eclectic form of the present investigation.

Now to summarize the various strands of the argument. Conceptually, I did not delve into the problem of defining a 'theory of meaning', but rather adopted implicitly some fairly intuitive notion along the lines of e.g. Grice (1957), involving both an aspect of *speaker's meaning* ('*mean* what you say') and *conventional meaning*

('what does this word *mean*'). Since both types of meaning require intentionality on the part of the language user, I agreed with critics of AI such as Dreyfus (1993) and Searle (1980) that artificial systems consisting of ingenious computer programs (either 'symbolic' or 'connectionist') which obviously lack intentionality, will necessarily lack meaning. Unlike Searle, however, I see no reason to assume that intentionality is a property of particular biological matter and therefore a robot with structurally similar body, forms of interaction and development to those of human beings could be a possible candidate for constituting an intentional system capable of meaning. This conjecture was supported through the Wittgenstein-inspired thought experiment of the 'deceased person-robot' (cf. Section 2).

The crucial difference between the position defended in this paper and most other 'robot-friendly' arguments is that it is not the least 'deflationist' with respect to critical (human) mental properties, in the manner of e.g. Dennett (1991, 1995). What I have tried to show is that the dilemma "Searle or Dennett" that most philosophical discussions concerning AI seem to deal with, is a false dilemma: we have the Vygotskyan alternative that *intentionality, self-consciousness and meaning are real emergent properties arising from the dialectical interaction between specific biological structures (embodiment) and culture (situatedness) through a specific history of development (epigenesis).* Since it is not inconceivable that the biological structures may be substituted with more or less isomorphic (and functionally equivalent) artificial structures, this line of reasoning leads to a positive answer to the question "Can a machine mean?".

Section 4 proceeded with the second question "How can this be achieved?" though instead of focusing on the constructive question – a much more technically-minded approach would be required for the task – it presented a schematic model of early human cognitive and linguistic development which is consistent with a multitude of empirical evidence, and (hopefully) internally coherent. That is, I focused on the empirical question, and only provided some hints at how the various forms of interaction and stage transitions may be realized by a robot. But there were some important implications for the *order* in which the major cognitive skills not only happen to appear, but *need* to appear. The central one can be summarized in the form of the following argument: linguistic meaning presupposes shared conventions, as a form of mutual knowledge. Conventions presuppose reflexive consciousness, allowing them to be learned and followed. Self-consciousness presupposes the perception of oneself as an intentional agent. Perception of oneself as an intentional being presupposes the perception of others similarly. Hence, other-intentionality, self-intentionality, self-consciousness and language form a possibly necessary developmental progression and an artificial system aiming at real – as opposed to simulated – language use would have to traverse it.

In sum, if the reasoning presented in this paper is in broad lines correct, then it represents an argument for the theoretical possibility of creating an artificial person which is quite different from traditional machine functionalism (Putnam, 1960). On the face of it, this claim is likely to cause rejoicing in robot-fans (consider

Commander Data from *Star Trek, II Generation*), and alarm in robot-foes (consider Arnold Schwarzenegger in *Terminator I*). It would seem, however, that the first group would be more justified in their reaction, since if language implies consciousness, and consciousness implies socialization – then the 'meaning machine' that we one day may create would be morally no better and no worse than its creators – which, of course, could be the crux of the problem but then it is not technology that is to blame.

But in truth, both reactions would be heavily exaggerated, since the *theoretical* possibility outlined in this paper may turn out to be *practically* impossible for a variety of reasons, such as the inability to build an 'artificial brain' that is both structured enough and flexible in the right way to allow the 'reprogramming' imposed on it by culture (for which human brains are certainly preadapted). Nevertheless, the fact that the possibility exists is tantalizing enough, and since its realization would amount to a radical breakthrough in our understanding of ourselves, it is bound to continue exercising the imagination of science fiction fans and cognitive scientists alike.

Acknowledgements

This paper was written as a result of the experience of spending the year 1999 as a post-doc researcher at the Department of Cognitive Science, Lund University (LUCS), and would have been impossible if it were not for the interactions with its members, both students and staff. I must single out, however, the contributions of Lars Kopp and Peter Gärdenfors. I also wish to thank an anonymous referee for constructive comments. While revising the paper during the year 2000 I was supported by a grant from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT).

Notes

[1]If the intention is "read into" the child's behavior by the parent himself on a particular occasion, this would constitute a step in the child's development, in the sense that it would help the child to *internalize* (cf. Vygotsky, 1978) the correlation between its own behavior and the parent's response and utilize this in a future occasion (cf. Section 4.2.3).

[2]Which as Searle (1995) (correctly) states, already "concedes too much to Strong AI" syntactic operations, or computation in general, is not a process which occurs in nature, independently of an interpreter. Thus computers can not even be said to be "computing", if it were not for us to interpret the operation of their electrical circuits in this way.

[3]However, cf. Rapaport (1988) for a more elaborate defense of the claim that syntax suffices for semantics.

[4]This thought experiment was initially suggested to me by Esa Itkonen in a discussion in Turku in 1996.

[5]Some examples of *wrong* combinations could be the following: The "closet-child" Genie (Curtiss, 1977) – adequate embodiment, while gravely deficient situatedness (isolated in a room, strapped to a potty, not exposed to any language) and development (this extreme depravation went on until she was over 13 years old). Inadequate embodiment: the chimp Viki was raised in a human like environment,

but because of obvious bodily deficits in vocalization did not acquire more than a few words (Hayes, 1951). The chimp Sarah (Premack, 1976) was superficially more successful in using plastic symbols, but because of a reinforcement learning approach, it is doubtful if these were really symbols for her rather than obstacles in the way of the reward. So far the most successful attempt of teaching meaningful language to a non-human has been in the case of the bonobo Kanzi (Savage-Rumbaugh et al., 1998) where a relatively good combination of the three features has been provided: The bonobos appear closer to us biologically than any other species (embodiment), Kanzi acquired many linguistic skills during the sensitive period while accompanying his stepmother Matata (development) and he was raised in a world in which language use was a natural part (situatedness), rather than being trained into it through conditioning.

[6]It appears to be a sound principle *not to over-interpret* infant, animal and machine behavior, and to opt for explanations based on mental states only when more simple mechanisms are inadequate. However, adequacy should be decided not just *locally*, i.e. with respect to a specific ability, but *globally* – with respect to the skills of the organism as a whole, and their development through time towards maturity.

[7]cf. Deacon (1997: 226) "The crucial role of the prefrontal cortex is primarily in the construction of the distributed mnemonic architecture that supports symbolic reference, not in the storage and retrieval of symbols. This is not just a process confined to language learning."

[8]By this I mean that the *actual* mapping need not (and given considerations of plasticity *should* not) be innate, but rather established as a result of experience. The *capacity* for such a mapping, however, e.g. in the form of reentrant mappings between topographic maps corresponding to the different modalities (cf. Edelman, 1992) should be provided.

[9]At least during the hypothetical 'sensorimotor period' lasting according to Piaget from birth until 18 months – though subsequent research has strongly questioned both the duration and the developmental structure of this hypothetical period.

[10]Though cf. (Neisser, 1995: 8): "Once an affordance has been perceived and the appropriate action initiated, that action must be appropriately controlled. In the view of most contemporary action theorists, such control depends in part on motor programs or schemata. Although Gibson did not share this view . . . it seems inescapable to me."

[11]It is not altogether clear what Edelman means in this citation, I presume that he has something like the following in mind: there is no causal relation between the boxes lying around in the room and the bananas hanging from the ceiling. Nevertheless, the chimpanzee perceives a (structural) similarity between the whole scene and scenes "in the wild" with tree branches and stones affording climbing, and after some experimentation with the boxes, manages to line them upon each other and thus to reach the bananas.

[12]cf. ". . . direct face-to-face interaction establishes a preconceptual form of knowing: knowledge of the "other and of the self as engaged with that other. . . . Ecological knowledge, obtained through interaction with the physical environment is equally direct. Both forms of perception are present from early infancy, long before the more sophisticated conceptual forms of self-knowledge begin to appear" (Neisser, 1995: 13).

[13]This hypothesis is not widely acknowledged, but the failure of both empiricist and rationalist attempts to explain how children acquire language, where consciousness of self and other doesn't play any role at all, is indicative.

[14]Human children pass the mirror test somewhat later than what would be expected given the 're-volution' at 9 months – between 15 and 21 months (Lewis and Brooks-Gunn, 1979). That could be taken either to support the position that the discovery of intentionality precedes the construction of a self-concept, or as an indication of more specific difficulties with understanding mirrors.

[15]This is, of course, a large claim which I am here only suggesting and not in position to argue for. For classifications and overviews of theories of L1 acquisition and development, cf. Ingram (1989), Bloom (1993) and Zlatev (1997, Chapter 5).

References

Baldwin, J. (1913), *Social and Ethical Interpretations in Mental Development. A Study in Social Psychology*, 5th ed. New York: MacMillan Press.

Balkenius, C. (1995), *Natural Intelligence in Artificial Creatures*, Lund University Cognitive Studies 37.

Balkenius, C. (forthcoming), 'Attention, Habituation and Conditioning: Towards a Computational Model', *Cognitive Science Quarterly* 1.

Bates, E. (1979), *The Emergence of Symbols. Cognition and Communication in Infancy.* New York: Academic Press.

Bates, B. (in press), 'On the Nature and Nurture of Language', in E. Bizzi, P. Calissano, and V. Volterra, eds., *Frontiere della Biologia The Brain of Homo sapiens*, Rome: Giovanni Trecani.

Bateson, M. (1975), 'Mother-infant Exchanges: The Epigenesis of Conversational interaction' in D. Aaronson and R. Rieber, eds., *Developmental Psycholinguistics and Communication Disorders. Annals of the New York Academy of Sciences* (Vol. 263). New York: New York Academy of Sciences.

Bloom, P. (1993), 'Language Development', *Handbook of Psycholinguistics*, New York: Academic Press.

Brooks, R., Breazeal, C., Marjanovic, M, Scassellati, B. and Williamson, M. (1999), 'The Cog Project: Building a Humanoid Robot', in C.L. Nehaniv, ed., *Computation for Metaphors, Analogy, and Agents*, Springer-Verlag Lecture Notes in Computer Science, 1562.

Bruner, J. (1983), *Child's Talk*, New York: Norton.

Bruner, J. (1986), *Actual Minds, Possible Worlds*, Cambridge, Mass.: Harvard University Press.

Buford, T. (1971), *Essays on Other Minds*, University of Illinois Press.

Butterworth, G. (1991), 'The Ontogeny and Philogeny of Joint Visual Attention', in A. Whiten, ed., *Natural Theories of Mind*. London: Basil Blackwell.

Butterworth, G. (1995), 'An Ecological Perspective on the Origins of the Self' in J.L. Bermudez, A. Marcel, and N. Eilan, eds., *The Body and the Self*, Cambridge, Mass.: MIT Press.

Carruthers, P. (1996), *Language, Thought and Consciousness*, Cambridge: Cambridge University Press.

Carruthers, P. (1998), *Mind and Language* 13(4), 457–476.

Changeux, J. (1985), *Neuronal Man*, New York: Pantheon Books.

Chomsky, N. (1980), *Rules and Representations*, New York: Columbia University Press.

Churchland, P.M. and Churchland, P.S. (1990), 'Could a Machine Think?' *Scientific American*, January 1990: 26–31.

Clark, H.H. and Marshall, C.R. (1981) 'Definite Reference and Mutual Knowledge', in Joshi, Webber and Sag, eds., *Elements of Discourse Understanding*, Cambridge: Cambridge University Press.

Curtiss, S. (1977), *Genie: A Psycholinguistic Study of a Modern-day "Wild Child"*. London: Academic Press.

Deacon, T. (1997), *The Symbolic Species: The Co-evolution of Language and the Brain*, New York: Norton.

Dennett, D. (1987), *The Intentional Stance*, Cambridge, Mass.: MIT Press.

Dennett, D. (1991), *Consciousness Explained*, Boston: Little, Brown.

Dennett, D. (1995), 'Cog: Steps Towards Consciousness in Robots' in T. Metzinger, ed., *Conscious Erperience*. Lawrence, Kansas: Allen Press.

Dennett, D. (1996), *Kinds of Minds. Towards an Understanding of Consciousness*, London: Phenix.

Donald, M. (1991), *Origins of the Modern Mind. Three Stages in the Evolution of Culture and Cognition*, Cambridge, Mass.: Harvard University Press.

Drescher, G. (1991), *Made-up Minds. A Constructivist Approach to Artificial Intelligence*, Cambridge, Mass.: MIT Press.

Dreyfus, H. (1991), *Being-in-the-world. A Commentary on Heidegger's "Being and Time, Division I"*, Cambridge, Mass.: MIT Press.
Dreyfus, H. (1993) [1972], *What Computers (Still) Can't Do. A Critique of Artificial Reason*, Third revised edition. Cambridge, Mass.: MIT Press.
Dunbar, R. (1996), *Grooming, Gossip and the Evolution of Language*, London: Faber and Faber.
Edelman, G. (1989), *The Remembered Present. A Biological Theory of Consciousness.* New York: Basic Books.
Edelman, G. (1992), *Bright Air, Brilliant Fire. On the Matter of the Mind*, New York: Basic Books, HarperCollins Publications.
Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D. and Pethick, S.J. (1994), 'Variability in Early Communicative Development'. *Monographs of the Society for Research in Child Developjnent, Serial No. 242. Vol. 59, No. 5.*
Firth, U. (1989), 'A New Look at Language and Communication in Autism', *British Journal of Disorders of Communication* 24, 123–150.
Fodor, J.A. (1983), *The Modularity of Mind*, Cambridge, Mass.: MIT Press.
Gallup, G. (1970), 'Chimpanzees: Self recognition', *Science* 167, 86–87.
Gibson, J.J. (1979), *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.
Gómez, J. (1994), 'Mutual Awareness in Primate Communication: A Gricean approach', in S. Parker, R. Mitchel and M. Boccia, eds., *Self-awareness in Animals and Humans*. Cambridge: Cambridge University Press.
Gopnik. A., Choi, S. and Baumberger, T. (1996), 'Cross-linguistic Differences in Early Semantic and Cognitive Development'. *Cognitive Development* 11(2), 197–227.
Grice, P. (1957), 'Meaning', *Philosophical Review* 66, 377–388.
Gärdenfors, P. (2000), *Conceptual Spaces. The Geometry of Thought*, Cambridge, Mass.: MIT Press.
Haeckel, E. (1874), *The Evolution of Man: A Popular Exposition of the Principle Points of Human Ontogeny and Phylogeny*, New York: International Science Library.
Harnad, S. (1991), 'Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem', *Minds and Machines* 1, 43–54.
Harnad, S. (1993). 'Grounding Symbols in the Analog World with Neural Nets', *Think* 2(1), 12–78.
Harris, P. (1991), 'The Work of Imagination' in A. Whiten, ed., *Natural Theories of Mind*. London: Basil Blackwell.
Hayes, C. (1951), *The Ape in Our House*, New York: Harper.
Hebb, D.O. (1955), 'Drive and the C.N.S.' *Psychological Review* 62, 243–254.
Ingram, D. (1989), *First Language Acquisition. Method, Description, and Explanation.* New York: Cambridge University Press.
Itkonen, E. (1983), *Causality in Linguistic Theory. A Critical Investigation into the Philosophical and Methodological Foundations of 'Non-autonomous' Linguistics*. Bloomington: Indiana University Press.
Itkonen, E. (1997), 'The Social Ontology of Meaning', *SKY 1997 Yearbook of the Linguistic Association of Finland, 49–80*.
Jusezyk, P.W. (1997), *The Discovery of Spoken Language*, Cambridge, MA: MIT Press.
Kozima, H. (1999), 'Attention-Sharing, Behavior-Sharing and the Acquisition of Language'. Paper presented at the international symposium on "The ecology of language acquisition", January 1999, Amsterdam.
Leder, D. (1990), *The Absent Body*. Chicago: University of Chicago Press.
Lenat, D. and Feigenbaum, E. (1991), 'On the Thresholds of Knowledge', *Artificial Intelligence* 47(1–3).
Leneberg, E. (1980), 'Of Language Knowledge, Apes and Brains' in T. Sebeok and J. Sebeok, eds., *'Speaking of Apes'*. New York: Plenum Press.
Lewis, M. and J. Brooks-Gunn (1979), *Social Cognition and the Acquisition of the Self*. New York: Plenum. Liège, August 11–15, 1997.

Locke, J. (1993), 'Phases in the Child's Development of Language', *American Scientist* 82, 436–445.
Malcolm, N. (1971), 'Knowledge of Other Minds', in T. Buford, ed., *Essays on Other Minds*. University of Illinois Press.
Mead, G. (1934), *Mind, Self and Society*. Chicago: University of Chicago Press.
Meltzoff, A. and Moore, M. (1977), 'Imitation of Facial and Manual Gestures by Human Neonates', *Science* 198, 75–78.
Meltzoff, A. and Moore, M. (1995), 'Infants' Understanding of People and Things: From Body Imitation to Folk Psychology', in J.L. Bermudez, A. Marcel and N. Eilan, eds., *The Body and the Self*. Cambridge, Mass.: The MIT Press.
Müller, R. (1996), 'Innateness, Autonomy, Universality? Neurological Approaches to Language', *Behavioral and Brain Sciences* 19, 611–675.
Neisser, U. (1988), 'Five Kinds of Self-knowledge', *Philosophical Psychology* 1, 35–59.
Neisser, U. (1995). 'The Self Perceived', in U. Neisser, ed., *The Perceived Self. Ecological and lnterpersonal Sources of Self knowledge*. Cambridge: Cambridge University Press.
Piaget, J. (1953), *The Origin of Intelligence in the Child*. London: Routledge and Kegan Paul.
Pierce, C.S. (1931–35), *The Collected Papers of Charles Sanders Peirce*. Vols. 1–4. Cambridge, Mass.: Harvard University Press.
Pinker, S. (1994), *The Language Instinct*. New York: William Morrow.
Premack, D. (1976), *Intelligence in Ape and Man*. Hillsdale, NJ: Laurence Erlbaum.
Putnam, H. (1960), Minds and Machines, in S. Hook, ed., *Dimensions of Mind*. New York: New York University Press.
Rapaport, W. (1988), 'Syntactic Semantics: Foundations of computational natural language understanding', in H. Fetzer, ed., *Aspects of Artificial lntelligence*. Dordrecht: Kluwer.
Regier, T. (1996), *The Human Semantic Potential: Spatial Language and Constrained Connectionism*. Cambridge, Mass.: MIT Press.
Reid, T. (1912) [1785], 'Essays on the Intellectual Powers of Man', in B. Rand, ed., *The Classical Psychologists*. Cambridge: Riverside Press.
Rizzolatti, G. and Arbib, M. (1998), 'Language within our Grasp', *Trends in Neurosciences* 21, 188–194.
Ryle, G. (1949), *The Concept of Mind*. London: Hutchinson's University Library.
Saussure, F. de (1916), *Cours de Linguistigue Générale*. Paris: Payot.
Savage-Rumbaugh, S., Shanker, S. and Taylor, T. (1998), *Apes, Language and the Human Mind*. Oxford: Oxford University Press.
Scassellati, B. (in press), 'Investigating Models of Social Development Using a Humanoid Robot', Presented at the 1998 AAAI Fall Symposium "Robots and Biology: Developing Connections" in Orlando, Florida.
Schoenemann, P. (1999), 'Syntax as an Emergent Characteristic of the Evolution of Semantic Complexity', *Minds and Machines* 9(3), 309–346.
Searle, J. (1980), 'Minds, Brains and Programs'. *Behavioral and Brain Sciences* 3, 417–24.
Searle, J. (1995), 'The Mystery of Consciousness', *The New York Review of Books*, November 2, 1995: 61–66.
Tomasello, M. (1995), 'On the Interpersonal Origins of the Self-concept', in U. Neisser, ed., *The Perceived Self. Ecological and interpersonal Sources of Self-knowledge*. Cambridge: Cambridge University Press.
Tomasello, M. (1999), *The Cultural Origins of Human Cognition*, Cambridge, MA: Harvard University Press
Tomasello, M., Kdruger, A. and Ratner, H. (1993), 'Cultural Learning', *Behavioral and Brain Sciences* 16(3), 495–511.
Tulving, E. (1985), 'How Many Memory Systems are There?', *American Psychologist*, April 1985: 385–398.
Turing, A. (1950), 'Computing Machinery and Intelliegnce', *Mind* 59, 433–460.

von Glasersfeld, E. (1997), 'Anticipation and the Constructivist Theory of Cognition', Presented at *International Conference on Computing Anticinatory Systems*.

von Hofsten, C. (1989), 'Transition Mechanisms in Sensorimotor Development', in A. de Ribaupierre, ed., *Transition Mechanisms in Child Development*. Cambridge: Cambridge University Press.

Vygotsky, L. (1978), *Mind in Society. The Development of Higher Psychological Processes*. Cambridge, Mass.: Harvard University Press.

Vygotsky, L. (1986) [1934], *Thought and Language*. Cambridge, Mass.: MIT Press.

Wittgenstein, L. (1953), *Philosophical Investigations*. Oxford: Basil Blackwell.

Wittgenstein, L. (1969), *The Blue and Brown Books*. London: Basil Blackwell.

Zlatev, J. (1997), *Situated Embodiment. Studies in the Emergence of Spatial Meaning*. Ph.D. Thesis. Stockholm University, Stockholm: Gotab.

Zlatev, J. (2000), 'The Mimetic Origins of Self-Conciousness in **p**hylo-, **o**nto- and **r**obotogenesis', Paper presented at Third Asia-Pacific Conference on Simulated Evolution and Learning (SEAL2000), Nagoya, Japan.