Project acronym: SEDSU

Project title: Stages in the Evolution and Development of Sign Use

Instrument: Specific Targeted Research Project (Strep)
Nest-2003-Path-3: "What it means to be human?"

Deliverable 27

Overall Theoretical Summary (WP7)

Due date of deliverable: March 2008

Actual submission: March 30 2008

Start date of the project: April 1st, 2005 Duration: 36 months

Organisation name of lead contractor for this deliverable:

Lund University, Sweden

Project co-funded by the European commission with the Sixth Framework Programme (2002-2006)		
Dissemination level		
PU	Public	
PP	Restricted to other programme participants (including the commission services)	
RE	Restricted to a group specified by the consortium (including the commission services)	X
CO	Confidential, only for members of the consortium (including the commission service)	

Overall Theoretical Summary of the SEDSU Project 4

Overall Theoretical Summary of the SEDSU Project

Göran Sonesson & Jordan Zlatev*
Lund University

1 Introduction

The project *Stages in the Evolution and Development of Sign Use* (SEDSU) was initiated under the FP6 call "What it means to be human". The major goal was stated in the project application: *To explain the evolutionary and ontogenetic development of sign use, clarifying its role on cognition.* In pursuing this goal, and in line with the theme of the call, the project was guided by the following general hypothesis:

> It is not language per se, but an advanced ability to engage in sign use that constitutes the characteristic feature of human beings. In particular, this implies the ability to differentiate between the sign itself, be it gesture, picture, word or abstract symbol, and what it represents, i.e. the *sign function* (Piaget, 1945), and thus to use (the same) sign systems for both communication and cognition. (Zlatev et al. 2006: 379)

The SEDSU project thus falls within the paradigm of *cognitive semiotics*, "integrating methods and theories developed in the disciplines of cognitive science with methods and theories developed in semiotics and the humanities, with the ultimate aim of providing new insights into the realm of human meaning production" (Cognitive Semiotics, Fall 2007: Editorial Preface, p.5). Semiotics: the general study of meaning (Sonesson 1989), theoretical linguistics: the study of the nature of language, and philosophy have provided concepts, coherence and a general perspective to our task. At the same time, empirical investigations involving methods from cognitive and developmental psychology, primatology, field and typological linguistics, and gesture studies have given us a wealth of new data, using four types of comparisons, in some cases in combination:

- **Cross-species**: between different species (capuchins, baboons, orang-utans, gorillas, chimpanzees and bonobos – as well as human beings); results may show differences in capacity, revealing either species-specific or species-typical characteristics.
- **Cross-cultural:** between different cultures, e.g. Himba vs. UK, Sweden vs. Thailand, UK vs. India, Amondawa (a traditional culture in Amazonia) vs. modern industrialised societies; results may show which aspects under study *could* be due to universal human characteristics, and which are due to culture, and in the latter case, indicate the nature of culture-specific semiosis.[1]

* We have conducted this report on the basis of the SEDSU Deliverables and publications written by the many people who have contributed to this large collaborative EU-project. Therefore, we need to gratefully acknowledge the following researchers, who at the same time should not be held responsible for our interpretations, and possible misinterpretations of their work: Jules Davidoff, Pam Heaton, Elisabeth Fonteneau, Amanda Ludlow, Julie Goldstein, and Steven Nugent (Goldsmiths, University of London), Chris Sinha, Vasu Reddy, Vera da Silva Sinha, Jörg Zinken, Kerry Hicks, Katja Liebal and Monia Conforti (University of Portsmouth), Wany Sampaio (Federal University of Rondônia), Srujana Jonnalagadda and Beena Chintalapuri (Osmania University, Hyderabad), Josep Call, Daniel Haun, Federica Amici, Alenka Hribar, Patricia Kanngieser and Bettina Blaesing (MPI-EVA, Leipzig), Mats Andrén, Ingar Brinck, Peter Gärdenfors, Mathias Osvath, Johan Blomberg, Soraya Osathanonda and Simone Löhndorf (Lund University), Caroline David (University of Montpellier 3), Giovanna Spinozzi, Elsa Addessi, Elisabetta Visalberghi, Lara Crescimbene, Alessandra Mancini, Valentina Truppa, Tommaso Stegagno, Giulia Castorina, Patrizia Potì, Martina Saporiti (ISTC-CNR, Rome), Carlo De Lillo (University of Leicester), Alessandra Amiconi (University of"La Sapienza"), Joël Fagot, Christine Deruelle, Cecilie Rondan, Bénédicte Hubert, Carole Parron, David Da Fonseca, Delphine Bastard-Rosset, Kate O' Konnor (CNRS, Marseille), Marc Martin, Seregaza Zohra and Jean-Philippe. Thank you!

[1] In this report, the term *semiosis* will be used to refer to all ways in which a phenomenon (e.g. a percept, an

- **Cross-populational:** between different populations within a culture: typically vs. atypically developing children; results can both help explain the nature of impairments such as autism, and contribute to the understanding of the interplay between endogenous and exogenous factors in the development of semiosis.
- **Developmental**: between different age groups in a single population; results can contribute to understanding the interaction between endogenous and exogenous factors in development (as in the cross-populational studies), and at the same time give evidence for stage-like transitions in the development of semiosis.

The results of these studies can be interpreted in various ways, but the goal of this theoretical summary is to furnish an interpretation which is consistent with the cognitive-semiotic concepts that have guided the project, and thus to bring the theoretical-empirical-theoretical spiral to a higher level.

The structure of this report will be organized in accordance with this conception. In section 2, we summarize our understanding of the most important cognitive-semiotic concepts for our research: (active) active perception, sign (function), icon, index, symbol, intersubjectivity, convention, imitation and stage.

On this basis, in Section 3, we formulate a general conceptual framework consisting of Four Levels of Meaning: *Active Perception, Imitation, Sign, and Convention*. Originally, SEDSU research was grouped around 5 different "domains": *Perception & Categorization, Iconicity & Pictures, Space & Metaphor, Imitation & Mimesis, and Intersubjectivity & Conventions*. While there is some overlap between this division and the Levels of Meaning framework, our investigation has lead us to view it as somewhat arbitrary in that it separated studies and results that belong to the same level, while sometimes placing together those that did not.

Hence, when in Section 4 we summarize and discuss most of the empirical studies within the project, we will do so in relation to the Four Levels of Meaning, and not to the five original domains. To anticipate, what these studies show is that there are semiotic differences between the usual case of comparison: Typically developing, Western, human individuals and the various comparison groups with respect to all four levels, but *the crucial difference that distinguishes human beings from other species concerns imitation, and its further transition into sign use.*

In Section 5, we take stock of these findings and turn to the professed goal of the project: a "coherent theory of semiotic development" (Zlatev et al 2006: 385). The most concerted such effort within the SEDSU project has been the Mimesis Hierarchy (MH) model (Zlatev 2007, in press). We claim that it presents a step forward compared to a similar theory (Donald 1991), which has been inspirational for us and which we also discuss in this section. It turns out that several of the results of the project can be interpreted in terms of the MH model. Others, however, require further interpretation and empirical investigation. In Section 6 we summarize the major conclusions of this theoretical summary and point out some directions for future research.

---

action, an imitation, a sign, an artefact, etc.) may carry meaning. Further on in the text, we will single out the *sign function* as a particular form of semiosis.

2 Basic cognitive-semiotic concepts

The concepts to be introduced in this section are not only those contained in the very name of the project, but also several others which are implied by those concepts, most of them also present in the description of the domains in the project description, though the purport of several of them have changed in part during our research. These concepts will be presented in two parts: first, some concepts that are traditional semiotic concepts, though they will be given a considerably more precise definition here. In the second place, a number of concepts which are more familiar from psychology, philosophy of language, and cognitive science, but which will be shown to be semiotic, in the more general sense of involving the production and interpretation of meaning.

2.1 The sign function and its varieties

Most semiotic theories do not define the concept of sign, or define it so vaguely, that it may be applied to just about anything. Here we first a propose a delimitation of the concept of sign, contrasted with other meanings (on which we will have more to say in the next main subsection, 2.2.1.), and we go on to define the relations of iconicity, indexicality, and symbolicity, as well as the corresponding signs.

2.1.1 Signs and pre-sign meanings

A *sign* is a kind of meaning, but not all meanings are signs. Perception is clearly meaningful to animals and newborn infants alike, but the capacity for *sign use* is a much more exclusive property. It is precisely because the capacity for using signs may be expected to distinguish stages in evolution and development, that it is important to separate signs from other meanings.

The sign is a meaning that is made up of two parts, traditionally known as *expression* and *content*. That the sign consists of two parts implies that the parts are separated. In Piaget's (1945; 1967a, b; 1970) terms, they are "differentiated from the point of view of the subject". This is not to say that the differentiation is "subjective", in the ordinary language sense. Indeed, in most cases, the differentiation is part of what is learnt by the child growing into a particular culture, i.e. it is *intersubjective* (see 2.2.2). However, what is differentiated within the sign may or may not consist of several objects (in the "objective" common sense world). Contrary to what Piaget suggests, a thing that is immediately continuous to another or which is a part of another in the common sense world may very well be differentiated within the sign (cf. Sonesson 1989; 1992; in press). In a parallel fashion, things that are similar to each other can be differentiated within the sign. Thus, there can be indexical (contiguity-based) and iconic (similarity-based), as well as symbolic (convention-based) signs.[2] If I see a branch sticking up over the house and conclude that there is a tree behind the house, this is a mere indexicality; but the marks on the ground left by the animal are indexical signs, clearly separated from the (part of) the animals having produced them. And the photographic print of my wife is clearly differentiated from my wife seen into the picture.

Indeed, a further differentiation may have to be made for certain purposes. The marks on the ground tell me "an elk was here before", and this is something distinct from the marks, as well from the elk, which is now somewhere is. Similarly, the colour configuration on the photograph is distinct from the perceptual impression of my wife, by even the latter is here now with me, while my wife is at home in Malmö. This is why we really have to separate

[2] For readers with a semiotic background, we would like to point out that we are well aware of this not being the Peircean definition of symbols, but it seems to us more coherent within a threesome of signs. Similarly, expression, content, and referent are of course not the members of the Peircean triad.

three parts of the sign, *expression, content*, and *referent*, where content is the standpoint taken on the referent by the sign user, as codified in some semiotic resource.

But differentiation is not a sufficient criterion. Each time we categorize two different things as belonging together (as opposed to categorical perception in which they are seen as instances of the same type), we must first differentiate the items put together. Still, categorization as such is not a form of sign use.

The sign is thus a whole made up of two parts, and therefore, there is necessarily some relationship between these parts. There is a *double asymmetric relationship* between expression and content. First, from the point of view of immediacy, expression is more accessible to consciousness than content. In the second place, content is more in focus (more prominent, more important) than expression. The founder of phenomenology Edmund Husserl (1939) formulated the definition of the sign (more precisely, "appresentation") more or less in these terms, but a similar view is implicit already in Augustine's conception of the sign (in our terms, the expression) as something which, by becoming conscious, makes us aware of something else (the content; cf. Deely 2001). This does not preclude other relations between expression and content being symmetric. It is common to suppose a substitutive relationship, which is a symmetric relation, between expression and content, but this may be misleading, since expressions are rarely used for the same purpose and in the same context as their contents.

It is clear from these descriptions ("subjective differentiation", "focus", "immediacy", etc.) that the sign is present in consciousness. This does not mean that it requires any high degree of awareness or any explicit effort of the will. But since it is present in consciousness, it must be possible to bring its constituency into awareness.

It will be noted that the present definition of the sign is considerably more specific than the one ordinarily employed in semiotic theory, notably by Peirce and his followers, for which all meaningful relations are signs, but it is at the same time much more general than the Saussurean notion, which tends to restrict the notion of sign to language and some other systems which are in some way similar to language. At the same time, it is more specific than both the Saussurean and the Peircean sign concepts, in that it clearly defines the requirement for two objects being called an expression and a content, while this is never done in the work of Peirce, and only by example by Saussure. The present notion of sign was first formulated by Sonesson (1989), taking his inspiration from the Piagetean idea of differentiation, and Husserl's definition of appresentation. Bates (1979:43) has also suggested that the sign (our expression) and its referent (i.e. the content) must be conceived as being similar and yet separate for a sign relationship to obtain. Daddesio (1995: 117) comments on this as follows:

> Given a physical mark (sound, movement, shape, etc.), *a*, and a particular class of things, *b*, that *a* is thought to stand for, let us consider three possible ways which organism can relate *a* and *b*. In the first instance, the organism fails to grasp any relation whatsoever between the two. /---/ In the first case, semiosis is thus absent. In the second case, the organism would be capable of relating the two, but instead of apprehending a relation between two distinct entities, it would simply react in the same fashion if presented *a* and if presented *b*. /---/ In the third case, the organism would recognize *a* and *b* as distinct but related."

From this it follows that it is impossible to conclude that if an individual only treats *a* and *b* as being distinct, that the particular relationship between *a* and *b* is necessarily one of appresentation (sign function). Daddesio's second case is that of categorization, which is important for perception. Given a prototype conception of categories, *a* and *b* may be treated as different just because they are differently central to the category of which they are perceived to form a part. Or they may be attended to differently, merely because one contains more, and more interesting perceptual properties, than the other. And, indeed, sign vehicles would tend to be "degraded stimuli", when compared to what they are signs of.

The problem of separating the expression and the content of a sign becomes particularly acute in the case of an iconic sign, in which, by definition, expression and content must share

at least some properties. Persson (2008: 10ff) has distinguished three modes of attending to pictures: *surface mode*, in which only "patterns, shapes, and colours, on the surface of the picture" are perceived (a picture of an apple is seen as patches of red and yellow); reality mode, in which the picture is seen as part of reality, instead of being about reality (the picture of an apple is seen as an apple); and, finally, pictorial mode, which involves both "an *expectation of separation*" and "an *expectation of likeness*" (the surface covered with patterns, shapes, and colours is seen a being *about* an apple). As Persson remarks, Fagot at al. (2000) independently made a similar distinction between picture processing in terms of "independence", "confusion", and "equivalence". *Mutatis mutandis*, the case of non-iconic signs is the same, though instead of an "expectation of likeness", there would be a more general "expectation of aboutness". It would be natural to think, however, that the expectation of separation (or rather: differentiation) cohabits more uneasily with an expectation of similarity than with the mere expectation of aboutness.

2.1.2 Iconicity and iconic signs (icons)

As so many other terms, "iconicity" is used in different ways in the literature. Here our intention is to follow Peirce, while avoiding the complications caused by his philosophical system, and taking into account some empirical counter-examples emerging from our work (Sonesson 1989; 1992; 1998; 2006; in press). From a strictly Peircean point of view, iconicity is really a list of properties, or more exactly, it is each one of the properties ascribed to an object. However, it is only when we have access to a least two objects, and we compare their properties, noting which are the same or different, that we have a relationship, which, with a Peircean term, may be called an iconic *ground*. It is thus the iconic ground that is of interest to us. From Peirce's works, it is not clear whether any two objects connected by an iconic ground should be said to form an iconic sign, or an icon, or whether they must also be connected by a sign relation. We will however suppose the latter to be true. Peirce is however clear about another thing: something being an icon does not preclude it from also being a symbol and/or an index. It therefore makes more sense to study the presence, in one and the same sign, of iconic, as well as symbolic and indexical, grounds.

The relationship binding together the two parts of the iconic ground is to repeat similarity, or sameness. There may be different types of iconicity. According to Peirce, iconicity can consist in simple properties, relations, or relations between relations (called, "images", "diagrams", and "metaphors", respectively). It is doubtful, however, that simple properties (in the sense of properties which are not relational in any sense) can be the cause of an iconic relation (though the impression caused by certain iconic signs such as the picture may be of an identity of properties). Pictures, which to Peirce were "images", certainly do not have simple properties in common with what they are pictures of.[3] Iconicity may also be more or less abstract or schematic, such as "upness" in Blissymbolics, or equilibrium, as the common factor of the balance and the law.

The kind of similarity involved is not necessarily visual. Indeed, it does not have to be perceptual in any obvious sense. Thus, for instance, Peirce claims that there is an iconic relation between Rumford and Franklin because both of them are Americans. But being American is not a property that is directly observable.

If an iconic ground is present in the relation between two objects (possibly together with indexical and symbolic grounds), and there is, in addition, a sign relationship between the same two objects, the whole made up of these two objects may be described as an iconic sign or an icon. Peirce claims that iconicity and the iconic ground (the possession of properties as well as their similarity) should be independent of the sign relationship, but this does not always seems to be the case, at least under some interpretations of independence. Thus, for

[3] Much more about the picture sign is said in Section 4.1, in relation to empirical studies attempting to attest it's understanding in infants and non-human primates.

instance, if somebody is painting the portrait of another person, then the similarity between the portrait and the person depicted is no mere coincidence, contrary to what may be true of the similarity between Rumford and Franklin. However, there is another, more important, respect in which iconicity may not always be independent of the sign function.

It is convenient to distinguish two kinds of iconic signs, which may be called *primary* and *secondary iconic signs*, respectively. In the case of a primary iconic sign, the perception of similarity is at least one of the reasons for positing a sign relation between two objects. This appears to be true about ordinary pictures. Secondary iconic signs, on the other hand, are signs in the case of which similarity is only discovered once they sign relation is known to obtain. Mallery, who studied the North American Indian Sign Language, says that many of the signs seemed reasonable once they were explained. That is, you see the similarity between expression and content once you are told what the content is. Thus, marking two descending lines on either side of the head may seem an arbitrary sign for woman. But when you are told it simulates the braids of the Indian woman, you discover the iconicity of the sign.

There are really two kinds of secondary iconic signs. Since there are two properties that are specific to iconic signs, iconicity between expression and content, and the sign function, either one of these may have to be discovered. When you see a car, you have no difficulty discovering its similarity to (other) cars. But you need to be familiar with the conventions of the motor-show, to understand that it is used as a sign. In the second case, as in Mallery's gesture signs, you know that what you see are signs, but you have to learn what they are signs of, in order to discover the iconicity.

2.1.3 Indexicality and Indexical signs (indices)

An index is a sign. It is thus made up of expression and content which are differentiated from each other, and while the first one is directly present to our awareness but not in focus, the latter is in focus and only indirectly present. In addition to its sign character, the index contains an indexical relationship between expression and content. Indexicality is a real connection, which, according to Peirce, exists independently of the sign relation. This means that the objects which serve as expression and content in the sign appear outside of the sign relationship, either in the neighbourhood of each other, or in such a way that the one of the items is a part of the other. These two kinds of indexicality can be called *contiguity* and *factoriality*, respectively (Sonesson 1989).

Just as in the case of icons, one may doubt, however, that indexical relationships are always independent of the sign relation. In one case, it is indeed because I am aware of a relationship of contiguity or factoriality between the objects serving as expression and content, that I am able to interpret the sign, and to interpret it in an indexical sense. The marks left by the elk on the ground tell me about the presence of the elk in this place at some earlier moment, because I am familiar with the kind of contiguity that may exist between the legs of the elk and the ground. The pointing finger is quit different. There is no pre-existing contiguity between expression and content. On the contrary, this relationship is created by the sign. There is simply no class of "pointed-out objects" independently of the sign. Thus, in the latter case, the indexicality of the sign is not independent of its sign character. Elsewhere, these signs have been called *abductive* and *performative* indices, respectively (Sonesson 1989). By analogy with iconicity, one may want to talk about primary and secondary indices, respectively, but it is not clear which one is which. The performative index appears to be analogous to the secondary iconical sign, because it is the sign that creates the indexicality and the iconicity, respectively. From this point of view, the abductive index is similar to the primary iconical sign, because indexicality and iconicity, respectively, is one of the reasons for seeing the sign as sign. However, is could also be argued that abductive indices and secondary iconical signs are more like each other, because both of them rely on some knowledge of relationships which is distinct from the sign relation.

2.1.4 Symbolic signs (symbols)

Like the icon and the index, the *symbol* is a sign. Unlike what happens in the case of those two other types of signs, the two parts making up the symbol are not connected by means of any other (relevant) relation than the sign relation. Symbols are therefore often said to be conventional (which is what Saussure and the tradition following him terms "arbitrary", but the associations of the latter term are often misleading; see 2.2.3). The convention in question may either by set up explicitly in given circumstances, or it may obtain implicitly in a particular society. Symbols may be part of complex sign systems, as is the case with language, or they may exist as single items. Though there is no necessary coincidence between these two divisions, the limitations of memory would seem to require that symbols set up explicitly should consist of relatively few items, not of entire sign systems, unless, of course, they are made up of written instructions which may be continuously consulted (as the book of Chinese signs used Searle's (1980) "Chinese Room"). On the other hand, those that have emerged spontaneously tend to form systems, since systematicity is both something in which symbols excel (compared to icons and indexes) and which may be a condition of their learnability in ontogeny (Deacon 1997, Zlatev 2003; Sonesson, in press). The relations between the terms discussed so far are shown in Figure 2.1.

There are numerous other usages of the term "symbol" that must be avoided here in order not to create confusion. In some traditions, the term symbol is used in the general sense in which sign is employed above. In other traditions, the term is used for a particular kind of iconic sign, whose iconic property is very abstract, though its expression is embodied in a concrete object. In this sense, Saussure described the balance as a symbol of the law. Although he referred to Saussure, Piaget employed the term in quite a different sense: for him it was a kind of meaning which was already differentiated, although expression and content remained "adherent". This could be understood to mean that they were not fully differentiated.

	Firstness	**Secondness**	**Thirdness**
Principle	Iconicity	—	—
Ground	Iconic ground	Indexicality = indexical ground	—
Sign	Iconic sign (icon)	Indexical sign (index)	Symbolicity = symbolic ground = symbolic sign (symbol)

Table 2.1 The relationship between principles, grounds, and signs, from the point of view of Peirce (as interpreted in Sonesson 2006).

In fact, Piaget tends to identify this with the relationship between the subject and what he/she is referring to. This becomes understandable if imitation is at the origin of the sign function. Even Peirce sometimes described the symbol as being based on habit, but this tends to make it border on the index. All these concepts are interesting, but they are better designated with other terms. We will in any case use the term symbol here exclusively to describe *signs based on conventions*.

2.2 Some further concepts of cognitive semiotics

So far, we have tried to demonstrate, against much of the semiotical literature, that the sign must be understood in a more specific, clearly delimited sense, if the concept should be useful, not only for the study of evolution and development, but also in the comparison of different semiotic resources. Especially from the first point of view, however, it is as important to show that there are other kinds of meaning (or semiosis), which somehow precede, and are a condition of possibility of, the sign function. We will start with *perception* itself, which is often treated as involving signs by many semioticians, but which we will consider as a more elementary form of semiosis, involving organization and categorization. To understood how perception may be an antecedent of the sign, however, it is necessary, not only to understand perception in close relation to action, but also to give an account of the process of attention, which is both part of perception and an ingredient of the sign function. *Imitation* (in one of the many senses of the term) may be that which mediates between active perception and the sign function. However, in order to function in communication, most signs also require some degree of *conventionality (normativity)*, which presupposes a capacity for intersubjectivity in the sign users. Finally, we will explicate the notion of stages, in development and evolution.

2.2.1 Active perception

From a semiotic point of view, perception is the primary level of meaning available to human beings, as well as to other animals. It is a level of meaning, because it supposes some kind of organization, and this organization may be different for different species, as well as for members of different cultures. In terms of the functional cycle defined by the biologist and precursor of biosemiotics, Jakob von Uexküll, each species has its *Umwelt*, the world as it is interpreted, made up of a *Merkwelt* and a correlated *Wirkwelt*, the actions which it may accomplish in relation to the environment; in addition, each human group or sub-group (and, as it appears, also some higher animals), have their own specific modification of such an *Umwelt*, which, with a term taken over from the philosopher Edmund Husserl, may be called their *Lebenswelt* (Lifeworld) their world-taken-for granted. As opposed to an objectively described ambient world, the *Umwelt* is characterised for a given subject in terms of the features which it perceives (*Merkwelt*) and the features which it impresses on it (*Wirkwelt*), which together form a functional circle (*Funktionskreis*). According to a by now classical example, the tick hangs motionless on a bush branch until it perceives the smell of butyric acid emitted by the skin glands of a mammal (*Merkzeichen*), which sends a message to its legs to let go (*Wirkzeichen*), so that it drops onto the mammal's body. This starts a new cycle, because the tactile cue of hitting the mammal's hair incites the tick to move around in order to find its host's skin. Finally, a third circle is initiated when the heat of the mammal's skin triggers the boring response allowing the tick to drink the blood of its host. Together, these different circles consisting of perceptual and operational cue bearers make up the interdependent wholes of the subject, corresponding to the organism, and the *Umwelt,* which is the world as it is defined for the subject in question.

An *Umwelt* or a *Lebenswelt* may work as a filter, leaving out everything which is not directly relevant to the biological processes characteristic of the species, or it may (and, in the case of a *Lebenswelt*, it will) function as a thematic device, which assigns different degrees of prominence to properties of the world without barring entirely access to any of them. Such levels of prominence may of course be constitutive of an *Umwelt* or a Lifeworld, or it may be produced on the fly, that is, in a given situation, by means of joint attention (see 2.2.2). It is useful to distinguish relevance from filtering, although they have something in common: the picking up a limited set of features from the totality of the environment. However, relevance, strictly speaking, does not exclude anything: it merely places some portions of the environment in the background, ready to serve for other purposes. Thus, in the case of

language, properties that are not relevant for determining the meaning of the words and the sentence still may serve to inform about the dialect, or even identify the person speaking In contrast, filtering simply crosses out that that is not let through the filtering device.

The difference between relevance and filtering no doubt has something to do with the capacity to be aware of the borders of one's *Umwelt*. It requires the ability to make the limits of the *Umwelt* into a theme. To the tick, to paraphrase Wittgenstein, the limits of its *Umwelt* are the limits of its world, but not so (in spite of Wittgenstein) to human beings. Or rather, the limits of our *Umwelt* are not the limits of our *Lebenswelt*. If the *Umwelt* is *a organised network of filters and/or relevancies*, it seems that development for the child consists in breaking out of one *Umwelt* and going on to another, broader one, until reaching the human *Lifeworld*. Between each *Umwelt* and the next, which encompasses it, there is always a "zone of proximal development" (Vygotsky 1978).

The loop going from perception to action and back again is already present in the correlation, postulated by von Uexküll, between *Merkwelt* and *Wirkwelt*. It is also prefigured in Piaget's notion of a sensori-motor stage of development. More directly, the concept of active perception derives from Gibson's (1962, 1979) "ecological" theory of perception and in particular his emphasis on perception as a form of activity. Gibson's insights have been developed and re-interpreted in multiple ways. According to Costall (2007), the most important characteristics of Gibson's concept of *meaning* is that it is neither equivalent to external sensory input, nor to "representations" generated in one's brain, but constitutes a dynamical, relational category that arises as an active perceiver interacts with an environment. This implies that perception requires self-movement (O'Regan & Noë 2001). For example, Gibson (1962) reports on experiments with blind subjects touching different shapes of cookie cutters. If the cutter was placed on the subjects' palms, they could only tell the correct shape with 50% accuracy. When the cutter was pushed randomly on one's palm, the subjects could identify it with 72% accuracy. Only by touching the cutter in a self-guided manner could they recognize the object in more than 95% of the cases. This study thus illustrates that perception is a form of exploration: it is not just a method to arrive at perception; rather perception is *equivalent* to the on-going exploratory process.

Perception confounded with action is also present in Gibson's notion of affordance, which is a kind of meaning distinct from reference, and thus from the kind of meaning conveyed by signs. There is no proper definition of the notion of affordance in Gibson's work, but he gives some suggestive examples: it is the graspability, or the edibility, of a thing. Graspability can be understood as the aptness to be grasped. Edibility must be interpreted as the susceptibility of being eaten. These are inferences which might be said using a phenomenological term, to be "sedimented" onto a object of the Lifeworld: accordingly, an apple, once it is seen to be an apple, is also perceived as something which may be grasped and then eaten, because these are events being known to have taken place (and "properly" so) with other apples at other times. Therefore, the apple is apt to be grasped and eaten, both in the sense of normalcy and normativity (Cf. Sonesson 1996a, etc.). While it is possible for graspability to be a property of things in some respect independent of culture, this could hardly be the case with edibility. Anthropological studies are full of examples of things being eaten in some places and considered entirely inedible in other places. And it is easy to think of other meanings that are clearly of the same kind as those mentioned and which are yet culturally specific. We just have to think about the dice. Suppose there is some human culture where die have not been invented: it might yet seem as if the throwability of the dice may be perceived directly by those coming from the proper culture. Similarly, for most people in contemporary Western culture, a computer keyboard has an immediate property of writability (not necessary less immediately present than the depressability of the keys). Thus, some affordances may be defined by our common *Umwelt*, other, "cultural affordances" (to coin a term which would be anathema to Gibson), derive from specific socio-cultural lifeworlds.

Another aspect of active perception is *social mirroring*. Meltzoff & Moore (1977, 1995)

have famously shown that neonates are able to copy a number of bodily actions such as tongue protrusion and mouth opening, practically from birth. More recently, such displays of "blind imitation" (the infant cannot see its own body part) have been shown for chimpanzees as well (Myowa-Yamakoshi et al. 2004). These results imply mappings between visual perception of body motion and kinaesthetic perception of one's own body and have been linked to specific neural mechanisms such as "mirror neurons" (e.g. Rizzolatti & Arbib 1998). While not offering evidence of innateness, the perception and classification of types of *biological motion* (Johansson 1973) through visually minimal information can also be explained as the result of mapping from vision to the *body-schema* (Gallagher 2005). Thus, perception of biological motion, as well as other higher-order structures (and even certain emotional states expressed by movement and posture) does not require to be thought about, or inferred, but is "perceived directly" in phenomenological terms (Gallagher 2005). This indeed brings us back to the insights of the Leipzig school branch of *Granzheitspsychologie,* contemporary with the more well-known Berlin school of Koffka and Köhler (cf. Sonesson 1989).

For the purpose of developing a theory of semiotic evolution, two issues thus become central: is human perception different also from those of non-human primates, and, if so, is this difference a prerequisite for the development of human specificity such as the emergence of sign use, or is it instead a retroactive effect on perception of the development of signs, or some specific semiotic resources such as language?

2.2.2 Intersubjectivity and joint attention

The concept of intersubjectivity has become highly prominent in interdisciplinary studies of the mind (e.g. Thompson, 2001; Bråten, 2007; Zlatev, Racine, Sinha, & Itkonen, 2008). In the most general term, it can be understood as *the sharing of experiential content (e.g. feelings, perceptions, thoughts, linguistic meanings) among a plurality of subjects*. Such "sharing" can take different forms, some more immediate, while others more mediated by higher cognitive processes, e.g. what Barresi and Moore (2008) call "understanding" as opposed to "simply sharing". Thus, intersubjectivity is not only more inclusive than constructs such as "theory of mind" and "mentalizing" (Baron-Cohen, 1995), but offers a qualitatively different perspective on social cognition. In brief, this perspective - with roots in the phenomenology of Husserl and his followers (Husserl, 1999 [1907]), the philosophy of Wittgenstein (Wittgenstein, 1953) and socio-cultural tradition in developmental psychology (Vygotsky, 1978) - may be summarized through the following statements:

- Human beings (and possibly some other "higher species") are primordially connected in their subjectivity, rather than functioning as "monads" who need to "infer" that others are also endowed with experiences and mentalities that are similar to their own.
- The sharing of experiences is not only, and not primarily, on a cognitive level, but also (and more basically) on the level of affect, perceptual processes and conative (action-oriented) engagements.
- Such sharing and understanding are based on embodied interaction (e.g. empathic perception, imitation, gesture and practical collaboration).
- Crucial cognitive capacities are initially social and interactional, and the specifics of human development are characterized by a process of gradual transfer of inter-personal to intra-personal modes of thought through a process known as "internalization" or "appropriation".

Different forms of intersubjectivity can be distinguished on the basis of the most prominent form of consciousness involved – perceptual, affective, or reflective (Honderich, 2006) – or on the basis of their intrinsic complexity (Bråten, 2007; Brinck, 2008; Zlatev, 2008). Zlatev, Brinck and Andrén (2008), focus on the phenomenon of *perceptual*

intersubjectivity, more commonly known under the label 'joint attention'. Though widespread, the latter term is rather ambiguous: sometimes it refers to the general case in which two or more subjects perceive the same target (Butterworth, 2003), while for others the term applies to more specific reciprocal states in which the subjects are also aware that they perceive the same target (Tomasello, 1999). Most often only *visual attention* has been described in the literature, but implicitly the descriptions have been thought to generalize to other modalities.

Building on previous work, Zlatev et al. (2008) define *perceptual intersubjectivity* (PI) as the process in which two or more subjects focus their attention on the same external target. Targets can be objects, events, spatial locations (e.g., a certain place to go to), or directions (e.g., a way in which to go). Two major types of PI episodes can be distinguished:

- Symmetric: when the target has already been noticed by both (or more) subjects. In discussing symmetric PI, we refer to the participants as subjects.
- Asymmetric: when the target initially is noticed by *only one* of the subjects (the 'sender'), and subsequently the other subject (the 'receiver') aligns her attention with the first subject's attention.

Furthermore, three general levels of PI may be distinguished, with symmetric and asymmetric counterparts.

The first level is synchronous PI (SPI), and consists in the subjects' simply *synchronizing their actions in time and space while performing similar individual actions* relative to a perceptual target. Synchronous PI is not an intentionally communicative behaviour, in the sense that the action is not performed with the goal of affecting the behaviour of another subject, though of course it may do so inadvertently.

The second level is coordinated PI (CPI) and consists of the subjects' *adjusting their actions* relative to a perceptual target. On this level the subjects' actions are intentionally calibrated in time and space and are *communicative* in the sense explicated above.

The third level is reciprocal PI (RPI), and is achieved by the subjects' *mutually matching their actions* relative to a perceptual target. On this level each action is intentionally adjusted in space and time to the actions of the other subject. Similarly to CPI, the action is communicative, but the interaction between the subjects is still more complex: Each subject will perform his or her actions in response to those performed by the other subject, with the result that the actions will be either *similar*, as in imitation, or *complementary*, as in turn-taking.

2.2.3 Conventions and common knowledge

The notion of *convention* is central, though controversial within the philosophy of language, linguistics and semiotics. One tradition, represented by Lewis (1969), Kripke (1982), Clark (1996) and Itkonen (2003) sees convention (and the closely related concept of *norm*) as a fundamental concept in accounting for the nature of language, semiotic systems, and social life in general. Another tradition, represented by Grice (1957), Davidson (1980), and Sperber and Wilson (1986), has been much more sceptical, arguing that "shared conventions" are not necessary for (linguistic) communication, but that rather a "passing theory" (which can differ for speaker and hearer) and open-ended interpretation would be sufficient: a project that has never been spelled out, and makes communication through sign systems such as language nearly impossible to explain. Finally, the term is used often in modern (cognitive) linguistics: "conventional metaphor" (Lakoff), "conventional imagery (Langacker), "conventional meaning" (Pinker), without any clear meaning. In his recent biophysical approach to language Lakoff has claimed that "Convention … makes sense only in neural terms" and "mirror neurons can be used to acquire such norms" (in a debate on Cogling List with Zlatev in

November 2002). But the first claim it itself difficult to make sense of, and the latter is doubtful, since a convention is a social phenomenon that cannot be in a single individual's head, cf. "the conventions of chess" or "the Geneva Convention".

A succinct definition of convention on the basis of suggestions made by the "pro-convention" tradition is the following: A convention is *(a) a regularity in behaviour, (b) that is commonly known in a community and (c) has an irreducibly normative nature* (i.e. members of the community expect other members to abide by it).

The notion of common knowledge implies reflexivity (conventions are known to be known). It is often criticized for implying an endless regress: to be sure that X is commonly known, I need to know X, know that you know X, know that you know that I know X etc. Itkonen (1978, 2003) has argued that this is not a problem, since for all practical purposes "three-level knowledge" is sufficient for a convention (or norm) to be commonly (or mutually) known. For example, to account for our common knowledge on which side of the road to drive, I must know the "right" side, know that you know the right side, and know that you know that I know the right side (for else I would worry every time I meet a car on the road.). However, such "knowledge" need not be propositional (explicitly formulated), for every case, due to (a) and (c): since I have observed the regularity of the behaviour in the past and since driving on the right is the correct side (depending, of course, on the conventions of the particular country) I can expect any other (mature, sane and sober) member of the community to drive on the right side, and to expect me to do as well. Thus, the normativity (as opposed to the "normalness" of behaviour patterns) of conventions can help resolve the "common knowledge problem".

Deflationist theories of convention are willing to accept (a), but reject (b) and (c) for being "too cognitive". But that reduces a convention to a habit or a stimulus-response association, which is far from being the same thing. Lewis (1969) wished to avoid (c), normativity, but this made him run into the problem of common knowledge, as suggested above. The most important kinds of conventions (for our purposes) are those that have the form of signs. To say that a sign is conventional (i.e. a symbol) is, according to the definition above, to say that (a) it is used by the community (as opposed to being "dead"), (b) the expression-content relation is commonly known and (c) there is a "right" way of using it (which of course, can change with time). Conventional signs in language include lexemes ('dog' means DOG) and grammar rules (constructions): "John kissed Mary" means KISSED(J, M). It is true that conventional signs are at least partly arbitrary, but "conventionality" and "arbitrariness" in the ordinary sense (as opposed to the technical sense of Saussure) are not synonymous, since most (linguistic) symbols are variously motivated by iconicity and/or indexicality (see 2.1). Index finger pointing is an example of a non-linguistic semiotic convention. Street signs are schematic pictorial conventions. Both are far from "arbitrary" but motivated by the nature of human bodies and brains.

2.2.4 Imitation

While phenomenologists have mostly linked intersubjectivity (see above) to bodily *perception*, psychologists have tended to emphasize *imitation* as a more active mode of understanding others, as well as of oneself. Baldwin (1894: 42) famously asserted: "My sense of myself grows by imitation of you and my sense of yourself grows in terms of myself", proposing a model according to which imitation serves as the primary means through which infants learn both about others, and develop self-consciousness through a dialectal process summarized by Zelazo and Lourenco (2003: 60) as follows:

> The dialectic starts with the presence of action that is (at least partially) outside of one's behavioural repertoire, and hence, viewed in terms of its outward or *projective* elements. By imitating, one comes to comprehend the *subjective* side of it; for example, one comes to

> appreciate the affect that accompanies it, or the effort involved. Once this happens, one tends automatically to eject this *subjectivity* back into the original behaviour.

Piaget (1951 [1945]) viewed imitation as the major source for deriving the first internal representations which he called "pre-concepts", suggesting the following developmental progression: (1) *sensorimotor imitation:* in which the model's action is imitated immediately following the stimulus; (2) *deferred imitation:* in which the performed action is a copy of an action removed in space and time, either of another, or of oneself; (3) *representational imitation:* in which "the interior image precedes the exterior gesture, which is thus a copy of an "internal model" that guarantees the connection between the real, but absent model, and the imitative reproduction of it." (Piaget 1951: 279). The pre-concepts that emerge through imitation can be said to constitute internalized, generalized imitations, which prior to being intra-personal have been inter-personal, in a way that is reminiscent of Vygotsky (1978).[4]

Gopnik and Meltzoff (1993: 352), otherwise proponents of a non-innate, gradually constructed "theory-theory" of mind, argue that "normal children are innately endowed with the capacity to imitate others. This provides a social bridge between the newborn and caregivers." Zlatev (2000) proposed that such an original "bridge" of pre-reflective intersubjectivity serves as the basis and necessary precondition not only for subjectivizing the Other, but of "objectifying" (i.e. explicitly conceptualizing) one's sense of self, moving from a largely unconscious *body schema* coordinating movements to a consciously accessible *body image* (cf. Gallagher 2005)

The term "imitation" is quite ambiguous in the literature, with some authors applying it to even the most simple acts of copying the motion of another, and others insisting on high-level cognitive processes such as the understanding of the model's goals/intentions. One way to resolve such disagreements is to regard imitation, like intersubjectivity, as a non-unitary phenomenon and, following Piaget (1951), to see it on several different evolutionary and developmental levels. One of the most basic forms of imitation, *neonatal mirroring* (Meltzoff and Moore 1978, 1983) does not require any explicit differentiation between the subject and the model. Thus, it corresponds to "primary" (Trevarthen 1979), or "proto-mimetic" (Zlatev, 2007) intersubjectivity, which requires neither (true) representations nor simulations (cf. Gallagher 2005, Chapter 9). Importantly, such imitation has also been demonstrated in chimpanzee neonates (Myowa-Yamakoshi 2004), while "true" imitation, involving the copying of both the *means* and the *goal* of *novel* actions, is something in which apes may have some basic abilities (Whiten et al. 2004), but which seems to be much more developed in our species (Donald 1991; Tomasello 1999; Arbib 2005). The relationship between imitation and intersubjectivity is a matter of current controversy: is understanding the "intentions" of others a prerequisite for imitation or vice versa (cf. Hurely and Chater 2005)? Using a broad, stage-based concept of imitation allows us rather to view the two as co-evolving (Zlatev 2008) and co-developing: "simple" forms of imitating the actions of others, overtly or covertly, allows for better understanding of their subjectivity and thus intentions. When children become experts in this, they also become better imitators, being able to distinguish between the means and goals of an action, and thus to engage in "true" imitation (Tomasello 1999).

It is furthermore by means of imitation that the "extension of conscious control into the domain of action" takes place according to Donald (2001: 261). Still, it must be noted that the act of imitation is in not equivalent to a sign (see 2.1). If I see somebody use a stone as a tool to crack open the shell of a nut, I may do the same thing, not to bring into mind the act of the other person I have observed, but to obtain the same effect. I attempt to realise the same act as

[4] At the same time, this typical sequence cannot be claimed to be a necessary invariant: sensorimotor imitation may be denied to children with motor disabilities, but they can nevertheless identify with others, and engage in covert imitation, i.e. imaginary re-enactment. Nevertheless, this would imply internalization (appropriation) of the inter-personal.

he did, that is, to open the shell up, so that I can take out the nut and eat it. Instead of producing an expression that is non-thematic but directly given which refers to a content that is thematic but indirectly given, I am realising a new instance of the category of acts consisting in cracking open a nutshell. Like in "emulation" (Tomasello 1999), I may of course try to obtain the same effect without attending to the adequate means, which would produce a failed act of imitation. Or, I may merely simulate the outer actions of cracking the shell open, without letting them have a sufficient impact on the physical environment, in which case I may either be engaged in symbolic play, play-acting, or simply practicing the movements.

Imitation may thus be said to be differentiated, in the sense of separating the mediator and that which is mediated, but it is not asymmetric, neither in the sense of focus, nor in that of directness. This also means that the purpose of the act of imitation is not to present the original act to another subject (or even to oneself). Bentele (1984) argued against Piaget that imitation does not manifest the sign function, but is a prerequisite for it: indeed, it will function as a sign only to the extent that it is taken to refer back to the imitated act, instead of just being another instance of the same kind. The same observation should apply to "symbolic" play, and was in fact made by Bentele in another context: the toy is a sign, he claims, only to the extent that the child takes it to represent the real thing, which cannot be true, for instance, in the case of a toy lion if the child has no experience of the real animal. In fact, the extent of the knowledge of the child may not be the relevant factor, but rather the attitude taken by the child: according to the degree of fictionality of the play world, i.e. its separateness from the real world, which is grasped by the child (cf. Winner 1982; Gardner & Wolf 1983) the lion may be made to instantiate a real lion act or to present into to the other children (Sonesson 1992b; 2007, in press).

Acts of imitation in this sense have two important properties: they are "public", in the very broad sense characterized by Donald (1991, 2001), i.e. they may be perceptually, often visually, inspected; and they can be copied by means of the observer's own body, with or without some additional implement such as a stone. In both these ways, imitation is different from episodic memory; and it is different from procedural memory in being a public record.[5]

2.2.5 Stages – in evolution and development

The concept of *stage* is fundamental in developmental psychology, and practically all (classic) theories on cognitive, emotional, and moral development from the 20th century have been "stage theories", despite their numerous differences, e.g. those of Montessori, Piaget, Kohlberg, Freud, Erikson, and Vygotsky. In language acquisition, "it is possibly the most often used term" (Ingram 1989: 32). During the last 20 years, however, the concept of "stage" has come under a good deal of critique for (a) being inconsistently defined (or not defined at all) (cf. Ingram 1989), (b) failing to predict the varying performance of and explain the competence of children in different cognitive domains (Gardner 1992), (c) being too discrete and static, while development is essentially continuous (Siegler 1996) and (d) often implying a complete replacement and "dismantling" of the previous stage, while "no emerging domain disappears; each remains active and interacts dynamically with all the others" (Stern 1998: xii). However, such critiques can be taken as implying the need to improve on the notion of *developmental stage*, rather than reject it.

In the study of human cognitive evolution, on the other hand, there seems to be an increasing use of the concept, with nearly all more elaborated recent theories implying some form of a stage model (Donald 1991; Deacon 1997; Johansson 2005; Arbib 2005). Those who are sceptical point out that a stage-model in evolution seems to imply qualitative transitions, while both the nature of evolutionary processes involving many small mutations and the fossil evidence seem to speak in favour of a prolonged gradual process. However, there is no contradiction between a stage-based theory and the evolutionary evidence. Firstly, it is

[5] For further justification of these claims, cf. Sonesson 2007, in press, and section 5.1. below.

possible to have relatively discrete transitions, or "punctuated equilibria" (Eldredge and Gould 1972) within a basically gradual framework, i.e. without this implying any form of "saltations" or sudden "macro-mutations" (Dessalles 2004). The likelihood for *fairly* rapid and abrupt changes is even higher when biological and cultural evolutions interact, as has most likely been the case with human cognitive evolution (Deacon 1997). Furthermore, the combination between paleological, neuroscientific and comparative psychological evidence seems to support qualitative differences (at least) between (1) monkey-ape common ancestor and apes (2) between the ape-human common ancestor and *Homo ergaster/erectus* and (3) between the latter and *Homo sapiens*. If we furthermore consider the *cultural* evolution that has proceeded since the stabilization of human biological evolution, we must also consider the difference between *cultures* without and with organism independent *representations* (Sonesson 2007) and between "mythic" and "theoretic" cultures (Donald 1991, 2001), or between "protolanguage" and language (Arbib 2005). The latter writes:

> When I speak of "stage" in phylogeny, I do not have in mind an all-or-none switch in the genotype that yields a discontinuous change in the phenotype, but rather the coalescence of a variety of changes that can be characterized as forming a global pattern that can emerge over the course of tens or even hundreds of millennia (: 123).

The question concerning the relationship between developmental and evolutionary stages is even more controversial. The existence of similarities between phylogenetic and ontogenetic development has been observed at least since Haeckel (1874) stated the famous "biological law" of *recapitulation*. At the same time the idea of such parallelism is often criticized for being aprioristic: why should the individual develop along a trajectory that is similar to that of the species, given that the temporal spans, the driving forces and the contexts of the two processes are so different (Sinha, 2002)? While these are valid objections, there is indeed a very good reason for what in broad lines can be described as a "recapitulation" of phylogeny: *the epigenetic nature of ontogeny* (Piaget, 1971; Sinha, 1988, 2002; Müller, 1996; Zlatev, 1997, 2001, 2003). This characterization implies that ontogeny proceeds through an ordered succession of stages that are guided, but not determined, by the genes, where "higher" stages require more time and experience than the "lower" ones. Since evolutionarily more ancient cognitive structures require less experience in order to mature than evolutionarily modern structures, the early stages of ontogeny will be dominated by evolutionarily more ancient mechanisms and the later stages by more modern mechanisms.

The "usefulness" of a more ancient neural/cognitive structure may in some cases be even greater, namely to serve the role of a "stepping stone" in the acquisition of the more recent structure, this role varying from one of facilitation to one of being a necessary prerequisite. In particular, there are good reasons to believe that a basic mechanism such as association learning constitutes a prerequisite for acquiring mimetic skill, while the latter is a prerequisite for the acquisition of language and other symbol systems (Donald 1991, 2001; Zlatev 2003), see Section 5. Finally, the SEDSU project deals, not with just any kind of stages, but stages of semiosis. Thus a succinct definition of *evolutionary and developmental stages*,would be: (a) Relatively stable periods, (b) in either evolution or development (without *presuming* any isomorphism), (c) characterized by the establishment of a novel semiotic capacity, and (d) in which previous stages and capacities are subsumed and viably integrated.

This is quite general, and different theories could be formulated depending on e.g. the nature of the *semiotic capacity* (which is a more general notion than sign use). The question to what extent each stage correlates with (and even better explains) behaviour in different cognitive domains (such as visual perception) is an empirical issue that should be addressed by a theory of semiotic evolution. Two such theories, that of Donald (1991, 2007) and the one developed within the SEDSU project, the MH-model (Zlatev 2007, 2008), will be discussed in Section 5, after first presenting a general descriptive semiotic framework, on the basis of the concepts discussed in this section (Section 3), and then summarizing most of the empirical results from the project (Section 4).

3 Four levels of meaning

Four basic levels of meaning may be deduced from the discussion of the basic concepts in the previous section. First there is the important distinction between perceptual (non-mediated) and sign-based (mediated) meaning. The first "level" includes various indexicalities and iconicities in (active) perception within the Lifeworld, which includes objects, situations and other individuals. As we have used the term, no sign use is *actively employed* on this level, though as stated in 2.1, there may be an *effect* of sign-use, without the perceivers being aware of this.

However, between these two levels we would wish to situate imitation, understood as the copying of novel acts, both on the level of means and (if applicable) ends. As described in 2.2.4, imitation occupies a natural intermediary position between perception (and action) and sign use by involving (a) displacement in time and (b) differentiation between means and ends. Finally, we can distinguish between non-symbolic (i.e. iconic and indexical) and symbolic (conventional) signs. In a "feature-based" analysis, this can be stated clearly as in Table 3.1.

	Displacement + Differentiation	Sign function	Convention
Active perception	-	-	-
Imitation	+	-	-
Non-conventional signs	+	+	-
Conventional signs	+	+	+

Table 3.1. The conceptual framework of Four Levels of Meaning

Note that this is *not* a stage-based model (cf. 2.2.5), constituting an empirical hypothesis of the way in which sign use emerges and develops, but a conceptual model that follows from the definitions of the concepts employed. In particular, it does not state that all forms of iconic signs, e.g. pictures, are "simpler", and thus emerge earlier, than symbolic (conventional) signs. The main purpose of this division of levels in the present context is to provide a structure for presenting and discussing the empirical results in the following section. As stated in the introduction, we believe that this allows a better meta-comparison than by holding to the five original "domains" and work-packages.

Having said this, we will see in Section 5 that this conceptual framework can to some extent map onto empirical theories of semiotic evolution and development. However, we need first to summarize and interpret the wealth of results provided by SEDSU during the past three years, to which we turn in the following (lenghthy) section.

4 Empirical studies and their interpretation[6]

In this section, we will attempt to summarize and interpret most of the results of the empirical studies conducted within the SEDSU project. In some of the cases, it has been necessary to sketch a background to these studies by discussing earlier research. We will use the Four Levels of Meaning conceptual framework from Section 3 for structuring the presentation: each of the first four subsections correspond to a particular level, and their order to the order of complexity: from perception, via imitation, non-conventional signs to conventions. The final sub-section will discuss two studies which do not belong to any particular level, since these deal with transitions between levels, and possibly in some cases: stages (Section 2.2.5).

Since we are writing this in the context of a "theoretical summary", it should not be surprising that, as we present the results, we use them to further develop some aspects of the conceptual framework, as well as attempt to bring into focus what we find most relevant for the efforts to develop theories of semiotic evolution and development (Section 5) and the final conclusions (Section 6).

4.1 Active perception

4.1.1 Local, global and configural processing

All relations in perception are basically iconic or indexical. Iconical grounds, when they do not form signs, give rise to categories (in the sense in which these are largely unconscious groupings of objects perceived, such in the case of "categorical perception"). Indexical relationships are either mere *contiguities*, or contiguities perceived to form part of independent objects, that is, *factoralities* (part-whole relationships). Within psychology, the importance of wholes to perception is well-known from Gestalt psychology onwards. While psychologists associated with the Berlin school, such as Köhler and Koffka, used the term "gestalt" to describe all kinds of wholes, some of their contemporaries in Germany proposed distinctions between different kinds of wholes (Cf. Sonesson 1989).

One such distinction that has become relevant for current research into human and animal perception, including the studies with the SEDSU project, is the distinction between *global* and *configural* processing in perception. Global processing focuses on higher-order features which tend to dominate lower-order features in such a way that the latter become blurred. Configural processing, on the other hand, yields gestalt-like properties which result from the interactions between the features within the whole. This distinction has been shown to be important for research on perception in autism, and has been defined in a similar manner. "Global processing may be considered the processing of the highest level of hierarchical stimuli. Configural processing may be defined as the processing of the interspatial relations between elements." (Rondan and Deruelle 2007: 199).[7] Both global and configural processing (and their respective features) are thus *holistic*. They can be opposed to

[6] References to the reports in the SEDSU project published as Deliverables, will be given in the form of a D, followed by the appropriate number. See the particular portion of the references.

[7] In the case where experiments concern pictures of faces configural processing may coincide with *resemanticisation*, described by Sonesson (1989) as characteristic of picture perception. The concept starts from the insight that there is nothing in the picture that resembles the segmentation of the sentence into words/morphemes or the division of words into phonemes/graphemes (the famous "double articulation" of language). Taken as separate lines and figures, all parts of the picture are meaningless. Once the whole is perceived as such, however, its meaning its distributed over its constituent parts: e.g., if the whole is perceived as a face, some parts of the whole will be seen as eyes, others as a nose, etc., which is quite different from what happens with the word "face" even when seen as a whole. However, such an identification of configural processing with the process of resemanticisation does not seem to be born out, since tests with geometrical shapes failed to produce different results from pictures of faces (cf. D11, Experiments 4-5).

local features, perceived independently of the whole of which they could potentially be members. Attention to local properties, in this sense, would be a clear form of local processing. It seems less clear whether "mere aggregates" (continuities which are not part of any higher whole), in the sense of Gestalt psychology, would also count as local.

It has long been assumed that human perception is characterised by a strong bias to process global prior to local information. However, some SEDSU findings appeared to challenge the assumption that global processing is a stable and universal characteristic of human information processing (cf. D6; D11; D12; D16). For example, in a study of *visual illusions* within the Himba, a traditional society in Namibia, found that while a sub-group did show a typical global processing bias on a classical local/global choice test, a pronounced local processing bias was observed in the majority of the individuals tested. Indeed, the reduced subceptibility of the Himba, as compared to British participants, to most of the classical visual illusions (Ebbinghaus, Müller-Lyer, but not the horizontal/vertical illusion) could be explained by local processing.

However, when the meaningless abstract shapes were replaced with (schematic) representations of culturally relevant objects (cows), the Himba appeared to experience a stronger illusory effect and failed to make more accurate judgements than the British controls. The findings suggested that cultural factors influenced the extent that the Himba bind stimuli from spatially separate parts of visual displays. As noted in D12, p. 5, however, "it may well be that the Himba did not interpret the figures as cows". This is really a question of picture perception and would have to be checked separately. Though this has not been studied, our own impression is that, to a non-Himba subject, the shapes may not even be convincing as "droodles" with the key given (See Figure 4.1).

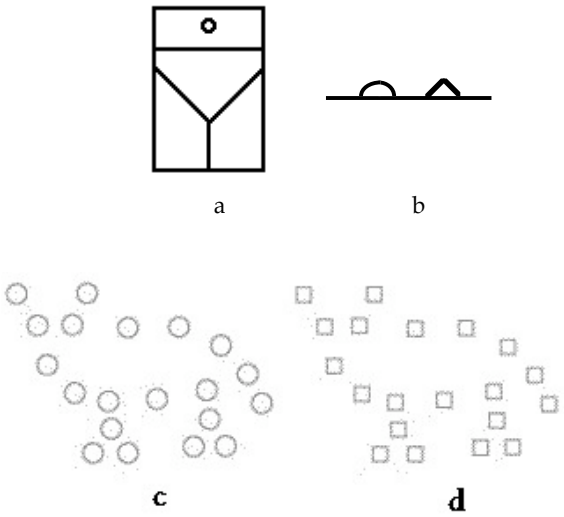


Figure 4.1. Droodles: a) Olive dropping into Martini glass or Close-up of girl in scanty bathing suit (from Arnheim 1969:92f). b) Carraci's key (Mason behind wall); c-d) the "schematic cow shapes" used in the experiment on the Himba

The droodle is a kind of limiting-case of a picture, which is mostly used in making jokes, inspired in the work of the Gestalt psychologist Rudolf Arnheim, who gives as an example a drawing that may show an olive dropping into a Martini glass or a close-up of girl in scanty bathing suit (from Arnheim 1969:92f). In this case, similarity is only discovered once we have been informed about the precise sign function, or when we have guessed at it, whereas in an ordinary picture the impression of similarity precedes the sign function. In the case of the presumed picture of a cow used in the tests, such a key would seem to be called for. However, as observed elsewhere (Sonesson 1994; 1996), drawings that appear to be as

rudimentary as the droodles shown here, turn out to be immediately identifiable as faces by human beings. Clearly, then, human beings need much less information to see a face – and of ecologically important phenomena of the human Lifeworld - than to see most other things. So it is conceivable that this information is sufficient for the Himba to see cows. Unfortunately, this has not been independently tested.

In order to further characterise the local bias seen in the Himba in the test with visual illusions, they were tested on a configural face processing task, using normal and "Thatcherized" versions of Himba and Western faces, as well as their inversions (D12). The results from this study failed to reveal significant differences between the Himba and British participants, except in the case of inverted faces. The combination of these results confirmed the need to distinguish between global and configural processing, when contrasting both with local processing: the visual illusion experiments showed that at least in some cases Himba subjects processed locally rather than globally (since they were not affected by the illusion); in the face interpretation studies on the other hand, they processed configurally rather than locally.

Is it then the case that "configural processing" is a human universal? A series of experiments carried out with children and adults with autism seemed to show that it was precisely configural processing that was impaired, rather the global processing (cf. D11). For example, in one experiment, ASD participants were presented with triads of figures, in which a composite target shape was shown at the top and two comparison shapes to the left and right below it. To contrast global vs. local bias in categorization, the target consisted of e.g. a circle made up of small squares and the alternatives, either a square made of small squares (local match) or a circle made of small circles (global match). The results were that the ASD group did not differ from controls in the number of global choices made. In a second version of this experiment, the intention was to contrast local and configural biases: here the targets where abstract face-like configurations or matrices of circles or squares. The choices involved either the same configuration, but with differences in the components, or an alternative configuration, in which the space between components was changed, leaving the components themselves unchanged. Here, interestingly, the ASD participants showed a clear bias for the local alternatives. Before concluding that impaired configural processing would provide the key to explaining the perceptual deficit in autism, though, it should be noted that even in the first version of the experiment, the "interspatial relations" between the elements making up large circle and the large square were changed, while those in the "global" choice, maintaining the overall figure remained the same. A contrast between the two versions that was not discussed in the deliverables, was that in the "circle" case, but not in the "face" etc. case, the choice could be biased by the difference in the global *concept* (e.g. circle vs. square) in the first, but not in the second case. So it could indeed be so that ASD is in general characterized by local bias (global deficit), save in cases where language-based or other higher-level concepts may influence categorization.

Another experiment revealed a global/configural deficit, but only for the processing of some sorts of information (D11, Experiment 2). Pictures of "hybrid faces" were composed by transposing a high-pass filtered image of e.g. a smiling woman and low-passed image of a frowning man and the tasks involved to identify the gender and emotion of the person in the picture. The results showed that while both the autism and control group preferred the low-pass (configural) stimulus when deciding the gender, the autism group relied more on the high-pass (local) stimulus in categorizing the identity and emotion of the face, in line with an explanation linking a perceptual processing deficit in autism with social-emotion factors.

These studies, as well as those discussed in Section 4.3.3 in relation to the processing of pictures, seemed to show that a simple explanation such as "an impairment in configural processing" cannot be the sole explanation for the perceptual deficit found in autism. A local bias was indeed found in some cases, a global in others and even (apparently) configural processing in third, depending on the nature of the task and the context.

Turning to non-human primates, there is more robust evidence that at least monkeys (capuchins and baboons in the SEDSU studies described in D8) are more proficient at processing the local features of visual stimuli than their global shapes. Two important findings from SEDSU research converge to demonstrate that baboons did not experience the so-called "Thatcher illusion" in the same way as human beings do: when the eyes and mouth of a human face are turned upside-down, as was first done with a photograph with the ex-British Prime Minister, the result in upright view is a grotesque face that is clearly different from the original. However, when inverted, the grotesqueness of the face seems to disappear

Figure 4.2. Illustration of the "Thatcher" illusion, these two apparently highly different stimuli will appear identical if they are turned upside-down

First, unlike humans, baboons presented with upright or inverted faces behaved in the same way. Second, while the baboons could promptly discriminate the face of two different individuals, they had serious difficulties to discriminate the normal faces from their "Thatcherized" versions, even when these faces were presented upright. Configural variations distinguishing the normal face from its Thatcherized version did not draw the attention of the baboons. The fact that there was no inversion effect, demonstrated that the canonical orientation of the pictorial faces had no special status for baboons, contrary to the case on humans. This may appear surprising, since there can be no doubt that faces are ecologically relevant stimuli also to baboons. A likely explanation of the results is that the baboons did not process the pictures as representations of real faces – that is, as pictures. (In terms of Persson's (2008) modes, this appears to be a case of surface perception of pictures.)

However, it should be noted that the study was not designed to investigate picture perception, but the perception of faces, which was taken to be equivalent in reality and through the mediation of pictures. A result such as this one may appear to throw doubt on the very premises of the study: if pictures may be used to manipulate the appearances of faces with human subjects, but not with baboons, then this is no doubt because to (some) baboons pictures of faces are not categorized as faces (in reality mode, as young children might employ), nor treated as a similiarity within a difference (in pictorial mode, as seems to be characteristic of adult, unimpaired human beings). It is also possible that it is the baboons' inability for configural visual processing that creates an obstacle to the perception of depicted faces in baboons, but other properties of pictures might also be involved, but this remains to be further tested. In fact, some of the studies of described in the following sub-section seemed to show at least some degree of configural processing.

In sum, while both global and configural perceptual processing can be said to be typical for human beings, they was shown to be non-universal, but rather depending on the nature of the stimuli, the task and the overall context. The Himba failed to show the expected global bias in the perception of illusions, but they did show a configural bias in the case of faces. People with autism seems to have a greater difficulty with configural than global (in general)

perception, but this too does not apply across the board. In terms of specificity, the studies with baboons did show a strong local bias in the processing of faces, but this could also be due to the pictorial nature of the stimuli. Also, as we shall see later (4.3.2), there were some results showing that chimpanzees (at least) can understand pictures in the pictorial mode, and although it seems clear that configural perception is a requirement for identifying pictures as such (as well, no doubt, for perception of picture according to the reality mode), we cannot so far distinguish this from global processing on this basis.

4.1.2 Perception of biological motion through point-light displays

Johansson (1973) presented human subjects with animation sequences consisting of few a dots, which were strategically placed on the joints of a person filmed in various activities. The participants promptly recognized the "biological object" (e.g., a person) depicted by the display and could even identify his/her action, confirming the contribution of motion cues to the identification of biological objects. Further research has shown that such highly degraded point-light displays (PLDs) suffice to derive a rich variety of conceptual information. Thus, typically developing human subjects can recognize the gender of the model or his/her identity (Cutting, 1978), the emotional connotation of the action (Dittrich, Troscianko, Lea & Morgan, 1996) and even the weight of an object being lifted (Runeson & Frykholm, 1981). It has also been shown that recognition performance is usually impeded when the dot-displays are presented upside-down (Pavlova & Sokolov, 2000), suggesting that the canonical orientation is a relevant dimension for this discrimination.

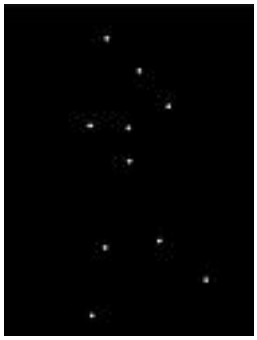


Figure 4.3. An example (a man running) of PLD used in the experiments

Two sets of experiments, with different setups and questions were performed by the group in Marseille, investigating the perception (and interpretation) of point-light displays (PLDs; see Figure 4.2), by (a) baboons and (b) three groups of impaired individuals. In the animal studies, 4 Guinea baboons previously exposed to (dynamic) picture-stimuli but not to PLDs were tested. A first experiment investigated their ability to discriminate between a PLD showing biological motion of a human figure walking, dancing of pushing an object, and "scrambled" versions of these, in which each individual dot had the same motion path, but their position was completely changed, so that the result was "biologically incoherent" (D13: 23). The results showed that the baboons required a very large number of trials to be able to discriminate between the two types of displays: 3556 for above chance performance, and 8372 to reach criterion of 80%. A second experiment showed that when the two types of displays were inverted, none of the four baboons showed a stronger reduction in performance for BM than non-BM displays. A third experiment "shifted" the positions of the lower and the higher parts of the displays, giving rise, in the case of the biological motion stimuli, to a situation in which the head and torso of the figure were below the feet, and correspondingly for the meaningless scrambled images. In this case there was essentially the same drop in performance in both cases, showing once again that the baboons did not perceive the biological (meaningful) and non-biological (meaningless) motion displays any differently.

Together, the experiments showed clearly that the baboons *did not see the displays as*

anything meaningful. However, since they could learn to discriminate between the BM and non-BM displays in the original, inverted, and shifted conditions, they must have done this by *remembering sub-configurations of dots*, and their associations with previously rewarded stimuli, rather than the motions of individual lights, which were identical. Thus, there was evidence for some form of configural processing, but not for the whole gestalt, and not based on the meaning of the pattern. It should be noted that, unlike in Johansson's original experiments, where the dots which formed moving patterns resulting from the movements of real biological beings, the stimuli were generated on computer screens, which is of course what allowed the sub-configurations to be scrambled and inverted. Considering the displays as pictorial signs (which they are spontaneously interpreted as by human subjects), we can say they were processed by the baboons in "surface mode" (Persson 2008), or Dadessio's "mode a" (cf. Section 2) — that is, they were clearly not signs. But even as purely perceptual stimuli, they showed something significant: that visual perception can be at least in a sense "configural", but still fail to deliver a meaningful perception.

The experiments with autistic adults, children and deaf children involving PLDs (D17) had different presuppositions and questions. Here, it was obvious that participants understood the displays as signs as soon as they understood the instructions "to watch the movies and to describe what was happening" (: 4). The issue was rather which of four different types of PLD representation would be most difficult to interpret: 10 actions (lifting, hopping…), 5 subjective states (itchy, bored…), 5 emotional states (surprised, frightened…) or 5 object movements (ball rotating, saw in action…). Both adults and children with autism had difficulty recognizing emotional and subjective states, but not the other two types, compared with matched typically developing controls. For the adults with autism, recognizing object motion was easier than the emotion/subjective conditions, while for the controls it was the most difficult condition. For the children, object motion recognition was the most difficult both for the autistic and the typical group, but for the first it was on the level of emotional state (both appr. 30% correct), while for the second, it was much less than the emotional and subjective state types of PLDs (both appr. 60% correct). In contrast, the experiment with deaf children showed that they performed as the controls. These results could be explained along with Hobson's (2004) theory of an impairment in "emotional identification", or Barresi and Moore (2008) proposal for "self-other matching" in autism. What they revealed also was that there did not seem to be any difficulty either with configural processing per se, or with understanding PLDs as pictorial signs in the impaired populations. The fact that the children with autism and those congenitally deaf were delayed in their acquisition of language, also suggested that differences in linguistic competence does not seem to have an effect on either visual perception or pictorial understanding.

4.1.3 Perception and categorization of colour

Another important aspect of perception that was also investigated within SEDSU was how it relates to categorization, in the sense of iconic ground (cf. Section 2.1). It is important to ask whether such categorization is influenced by the use of different languages, but also, more generally, by differentiated sign use. And even more generally, by the organization of meanings separating different cultural worlds.

The by now classical works of Berlin & Kay and of Rosch are commonly considered to have demonstrated that not only the color spectrum, but essential parts of the world of our common experience, tend not to be divided not "arbitrarily", but to follow certain common principles of segmentation. This was taken to deny any strong role of language for such categorization. Even if this were so, which part of the SEDSU research aimed to question, it does not imply that it is the world *as such* that possesses such a structure. Rather, in the terms of Uexküll it is the *Umwelt*, defined primarily by the sense organs the organism has access to (cf. Section 2.2.1). Thus, we should expect not only all human beings, but also other primates having similar sense organs, to be susceptible to the same perceptual segmentation and

categorization. However, to the extent that the neurobiological underpinnings responsible for colour perception can be identified with the relevant *Umwelt*, these are not sufficient to explain the difference in colour perception between human beings and other primates found in SEDSU research (cf. D12). There are also intriguing differences between at least British and Himba subjects, showing that the perception and categorization of colour cannot be a matter of a universal, biologically defined *Umwelt*.

The Himba language possesses only very broad colour terms and this could be correlated with some aspects of Himba speakers' categorization of colour. However, Franklin at al. (2005) argued that Himba toddlers sort colors into these same broad categories prior to the acquisition of colour terms, suggesting that language is not implicated in this aspect of cognitive organisation. However, this study tested categorisation around the blue-purple boundary, a region of the colour spectrum that the Himba associate with cattle terms. As this raised questions about the validity the findings, Himba toddlers were again tested, this time using stimuli that encompassed the blue-green and the blue-purple category boundaries. In the task the children were briefly presented with a colour stimulus, to be followed, after a short brief interval, by presentation of two other colour stimuli. The child was asked to indicate which of these was the same as the original. While the two comparison stimuli were equidistant from the original, some crossed category boundaries, and increased identification of these provided evidence for categorical perception. The findings from this study found no evidence for categorical perception in Himba toddlers. From these results, one can conclude that it is indeed possible that it is the colour terms that determine the colour experience of Himba adults (D12). However, another possible explanation would seem to be that it is the experience with cattle within the Himba lifeworld which determines both the colour terms and the categorization around the blue-purple boundary, and which thus also accounts for the different treatment of the blue-green boundary.

In order to further investigate the possible effect of language on Himba perception, categorical perception was tested in adults. The stimuli used in this experiment were pairs of linguistically distinct cattle types familiar to the Himba (the characteristic examples of which are shown in Figure 4.4), which were morphed in a continua of 10 stimuli, with the category boundary lying between numbers 5 and 6 (Figure 4.5). The results from the study showed a clear difference between the Himba and English controls who did not possess names for the cattle used in the experiment. Unfortunately, we cannot so far conclude that language is here the causal factor in categorical perception. In fact, it seems at least as reasonable to think that it is the experience that the Himba are having in their everyday lifeworld of two kinds of cattle which has led them to introduce the two terms. Indeed, the prototypical organization of the categories is hardly understandable if perception *cum* habitual action has not been the principle factor in the segmentation of the Himba cattle experience.

Figure 4.4. The two extreme poles of the cattle concept division

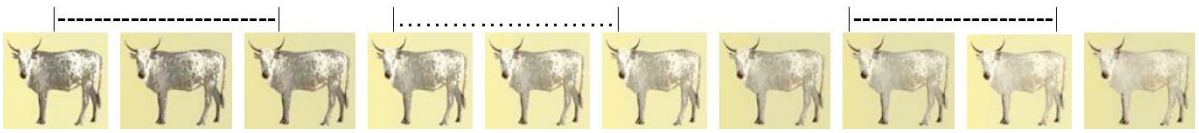


Figure 4.5. Intermediate morphs created as experiment material

4.1.4 The use of landmarks in non-human primates

The most active kind of perception studied within the SEDSU project concerned the task of localizing hidden food using surrounding objects. Landmarks play a crucial role in guiding an animal's search for goals such as food sources, nests or sleeping places. When confronted with an array of landmarks under experimental conditions, animals were shown to favour either of two strategies when locating a target: beaconing or vector search. Entities in the environment that are used as markers and that are contiguous with a target location are generally referred to as "beacons", whereas more distal entities are called "landmarks". Humans under the same circumstances applied the more abstract middle rule, said to involve the relationship between as least two landmarks and some object determined in relation to them. Since the latter is described by the authors as having to do with configural processing, beacons and landmarks must no doubt be taken to be some kind of local cues, rather in the sense of aggregates than of single features.

In order to investigate whether using the middle rule remains unique to humans, three studies were conducted with a closely related species: bonobos (D23). The experimental set-up consisted of a 47cm x 47 cm board with 81 holes and three different arrays of uniform landmarks that were presented in three subsequent studies. At first, the bonobos were always trained to find a reward in the centre of the array, which was relocated from trial to trial. Then, they were tested with an expanded array, by means of which their favoured search strategy could be determined. With respect to two of the expanded arrays - a four-landmark array with landmarks forming the corners of an imaginary square, and a horizontal two-landmark array with three holes between landmarks - the bonobos did not use the middle rule, but rather beaconing or vector search. However, when a horizontal array of two landmarks with five holes between landmarks was used, one bonobo preferably searched in the middle of the array. Our colleagues argue that this particular finding is due to the middle of the array coinciding with the middle of the board, in which case the application of the middle rule remains a strategy exclusively used by humans.

A series of similar experiments were also realised with capuchins (D18) concluding that there are some, very limited conditions, under which capuchins can learn the abstract middle rule and thus use landmarks configurally. When trained with a four-landmark configuration, capuchins searched in the vicinity of one or two landmarks, thus simply associating one or two landmarks to the food. Second, when presented with a two-landmark configuration, which is translated and rotated and only one inter-landmark distance is used, capuchins showed an elemental use of landmarks, at the maximum. However, when capuchins were trained with two inter-landmark distances and one orientation, at least one capuchin searched in the middle of the expanded configuration. Third, the use of a small-scale space seemed to be another necessary condition to elicit a complex use of landmarks by the capuchins. In conclusion, the number of elements in a configuration, the type of information received during training and the scale of space all affect capuchins' performance.[8] Compared to adult humans, capuchins are highly constrained in their access to configural spatial representations and abstract geometrical rules.

The conclusion seems to be that matter the exact extent to which apes and monkeys are able to use the middle rule, human spatial perception (at least in adults), is very different from that found in non-human primates, a least prior to special training.

4.1.5 Conclusions

SEDSU research into perception focused on the visual modality (though cf. e.g. D11, Exp 3), and involved comparison of human and non-human subjects, impaired (above all ASD) and

[8] In fact, a bigger quantity of elements were never tested with a small search space, so it cannot be known whether both these conditions, or any one of them, is necessary for the the middle rule to be applied.

typical individuals, Western and Himba cultures. The distinction between global and configural processing was important, since it could explain the lack of global processing in the Himba in the case of illusions, coinciding with configural processing of faces, and an inverse pattern in autism, with preserved global processing in some cases, and impaired configural processing. The method for using landmarks demonstrated by bonobos and capuchins show clearly that configural processing is much more difficult for them than for human beings. Nevertheless, this could not explain all results. Even baboons, which showed non-configural processing of faces, learned to generalize the patterns of point-light displays on the basis of "sub-configurations" of moving elements. And the clearest results concerning impaired perception in autism concerns social-emotional information (in PLDs and pictures of faces), rather than a general configural processing deficit.

A pattern that emerged is that the results of studies studying perception depend to a high extent on what the stimuli are, and how they are interpreted. If the "cows" in Figure 2.1 were indeed seen as such by the Himba subjects, that could explain why in this condition they were relatively influenced by the visual illusion, but not when what they saw were simply abstract patterns. We suggested also that we need to know if the baboons saw any faces (in either reality or pictorial modes) rather that patches of color in a "surface mode" before we can generalize that they do not perceive faces configurally/globally. Finally, a culture-specific Lifeworld, rather that a universal *Umwelt* was shown to be important for the perception and categorization of colour. However, we do not yet know whether these Lifeworld differenes are connected with sign use in general, or language in particular. To do so, signs need to be clearly demarcated from other kinds of meanings, and language needs to be disentagled from other kinds of signs.

4.2 Imitation – and its precursors

Chimpanzees, bonobos, gorillas and orangutans have been shown to be able to copy others' behaviour to some extent and in seemingly various ways. The literature on ape imitation in the broad sense of the term (see Whiten et al. 2004 for a review) involves a number of experimental paradigms: Neonatal mirroring, Instrumental tool-use, Experiments with "artificial fruits" (puzzle boxes) and "Do-as-I-do" tasks. Imitation may be said be attested when *an observer copies something of a model's bodily actions*, may it be in a laboratory tool-using situation, natural foraging situation or mimicking in a do-as-I-do game. Other forms of observational or social learning, like *emulation*, available in several versions (Byrne 1998; Custance, Whiten and Fredman 1999), *response facilitation* (see Byrne 1999), and *stimulus- and local enhancement* (see Tomasello 1996), can inform observers about properties of the environment and objects, or the connection between certain manipulations and outcomes, but do not involve copying bodily actions *per se* and thus it is not clear if they involve imitation.

The failures of great apes to fully copy behaviour in several early studies led Tomasello, Kruger and Ratner (1993) to conclude that apes emulate, but cannot imitate, since imitation in their view entails copying both of the goal and the method to bring about that goal, whereas for emulation, it is sufficient for the goal to be taken over from the model. For Tomasello (1999) imitation thus has a "theory of mind" component by definition, and it was initially believed that apes lack an understanding of others as intentional agents. Since then the field of primate imitation studies has expanded and the MPI Leipzig group has modified their views on apes' abilities to understand intentional agency (Tomasello et al. 2005). Recently, Buttleman, Carpenter, Call and Tomasello (2007) have even attested the ability of "rational imitation" (i.e. to copy an action when it is deemed essential for achieving a result, but not otherwise) in enculturated chimpanzees.

Thus, there is now a near-consensus that at least some apes can *to some degree* both imitate and understand the goals of others', both human beings and apes. What is not yet clear in both of these cases is "to which degree" this is true, and if there are differences between different species. SEDSU experiments testing for basic skills of intersubjectivity will be discussed in Section 4.4 as precursors of the ability to entertain conventions. Here, we summarize and briefly discuss a number of studies which explored ape capacities of (proto)imitation (cf. D20).

4.2.1 Spontaneous copying

Amichi and Call presented chimpanzees, gorillas, orangutans and bonobos with videotapes of conspecifics performing actions that were in the subjects' repertoire and measured whether subjects increased the frequency with which they executed those actions. The aim here was to assess the propensity of subjects from different great ape species to show *response facilitation*, that is spontaneously reproduce a number of different actions of conspecifics that were already in the subject's repertoire.

All species were comparable in the pretest in all behaviours except nose wiping for which chimpanzees and bonobos showed a higher rate. The results showed that chimpanzees yawned more after seeing videos of conspecific yawning and wiped their nose less often. Such effects were specific to those actions, e.g. the chimpanzees did not alter their scratching rate upon seeing other apes scratch. None of the other ape species showed any effects upon seeing videos. The result on spontaneous yawning in chimpanzees replicates the finding from another study (Anderson et al., 2005), and similar to that only about a third of the subjects showed the effect. However, it is conceivable that no effects were observed in the other species due to the small sample size.

4.2.2 Recognition of imitated actions

Haun and Call investigated whether great apes are capable of recognizing that others are imitating their actions. This research paradigm has been implemented successfully in human children and in one chimpanzee (e.g., Asendorpf, 2002; Nielsen et al., 2006). A comparison between the results of this recognition approach and the production approach previously described was aimed at investigating the relation between these two complementary skills.

The study was designed to test whether non-human great apes recognize when they are being imitated and as a result display testing behaviour similar to that found in human subjects in such cases. The results showed that apes (regardless of the species) spent significantly more time engaged in testing behaviours when the experimenter displayed contingent matching. A further study found no support for the possibility that something inadvertently peculiar about the experimenter's behaviour during the contingent/matching condition led subjects to respond more in this condition.

In combination with the study of spontaneous copying, the present study suggests that recognition of others' imitation is a more ancient evolutionary adaptation than some forms of response facilitation such as contagious yawning. This would appear to support Anderson et al's (2005) interpretation that the latter involves a form of identification with the model, and is not the result of a simple release mechanism.

4.2.3 Do-as-I do imitation

The most demanding imitation skill involved training a chimpanzee to imitate body movements on command and to test his ability to generalize to novel actions. Moreover, the researchers also presented videos, pictures and drawings of familiar actions to see if the chimpanzee would reproduce the actions depicted on them.

A 5-year-old nursery-reared chimpanzee (Alex) participated in this study. He was nursery reared until the age of three years together with two other chimpanzees. Before the beginning of the current study he had already received some training in the "Do as I do" task. In the context of play, the caretakers showed him actions and enticed him to repeat them in a role reversal situation. Alex also received some clicker training paired with the delivery of small food rewards for reproducing actions appropriately although this was not done systematically. At that time Alex was taught 12 actions such as "put hand on nose and blow air loudly through lips". Most of these actions were invented by Alex and the caretaker repeated them.

Alex was trained on 15 novel actions, along with 9 of the actions that he had learned to imitate earlier, shown in Table 4.1. After that, he was tested on 46 novel actions, the same as in Custance *et al.* (1995) study. It took Alex 92 sessions during 7 months to learn these 15 novel behaviours (although, he did not fully learn "touch armpit").

Table 4.1. Actions on which Alex was trained, including 9 which he could imitate prior to the study, and 15 novel actions.

Training actions	
Actions imitated prior to the study	
1. Clap:	the palms were hit together several times
2. Hit head:	the top of the head was hit with right hand several times
3. Fingers to cheeks	index fingers were pressed to both sides of the cheeks
4. Shake lip:	the lower lip was taken with thumb and index finger and then moved vigorously
5. Hit stomach:	the stomach was hit several times with right hand
6. Shake ear:	the right ear was taken with thumb and index finger and then moved vigorously
7. Smack:	a sound was made with a tongue
8. Protrude tongue:	the tongue was put out of the mouth
9. Fish face:	mouth was closed and the cheeks were suck in
Novel actions	
10. Shake hand:	the right hand was shaken loosely from the wrist
11. Raise one arm:	the right arm was put into the air
12. Stamp foot:	the right foot stamped the floor several times
13. Pat stomach:	the stomach was patted alternately with the palm of each hand several times
14. Raise two arms	both arms were put into the air simultaneously
15. Touch chin:	the right index finger was placed on the chin
16. Praying hands:	both palms were touching each other
17. Wipe face:	the palm of one hand was wiped down over the face several times
18. Slap floor:	the floor was slapped several times with the right palm
19. Raise foot:	the right foot was raised from the floor
20. Wipe floor:	the right palm was wiped from side to side across the floor several times
21. Touch armpit:	the left arm was raised and the right index finger was placed on the left armpit
22. Grab wrist:	the left wrist was grasped by the right hand
23. Swing arm:	the right arm was swing back and forth several times
24. Wipe hands:	the palms were wiped together several times

The results showed that Alex became quite proficient at matching the actions displayed by an experimenter. However, this ability was mostly restricted to familiar actions and it did not extend to most of the novel actions. He only reproduced about 17% of the novel actions in the transfer test, which is comparable to the results of a previous study with two chimpanzees (Custance et al., 1995). This result confirmed that reproducing novel actions, even after having received extensive training is a difficult task for chimpanzees.

The results could not be due to the fact that Alex was not motivated to imitate the novel actions, since he not only reproduced familiar actions with great accuracy but the pattern of his failed attempts (mismatches) strongly suggested that he was trying to reproduce the actions: Alex often looked intently at the novel actions without initially attempting to reproduce them whereas he responded quickly upon witnessing familiar actions. This suggests that he was motivated to copy all actions and, more importantly, that he detected even subtle differences between familiar and novel actions.

Since the novel actions were not more difficult to carry out, it was concluded that that the problem lies in the interface between observing and reproducing an action, that is: in translating visual information into motor commands for action execution. An analysis of the type of action, the body part that executed the action, and the place where the action was applied (in the case of transitive actions) found no differences between familiar and novel actions. The only difference observed was that, for novel actions, Alex produced less mismatches with transitive compared to intransitive actions. A similar result was found by Call (2001) in an orang-utan and a chimpanzee.

A further set of experiments presented Alex with videos, photographs and drawings

showing the actions that he was able to imitate. However, since these results are most significant for the ability to interpret signs, we discuss them at the end of the following section.

4.2.4 Imitation from complete and incomplete representations

In a later experiment (Call, Hribar, & Sonesson, in press) the imitation of a small range of actions were contrasted when solicited from (a) photographs of the final state of the action and (b) photographs of a pre-final state of the action.

The procedure was as follows: Alex and experimenter (E) were sitting opposite to each other with a panel between them. E had a laptop on her lap. Alex was shown an action and was asked to repeat the action. There were 3 conditions: 1) End state condition: Actions were presented on the laptop as still photos of the experimenter demonstrating the actions. Two sorts of actions were presented: 10 non-repetitive actions and 5 repetitive actions. All photos showed an end state of an action (Alex had already been tested with "Do as I do" procedure with most of the photos used in this condition). 2) Incomplete condition: Actions were again presented on the laptop as still photos of the experimenter demonstrating the actions. The same actions as in End state conditions were used, but in this condition the photos showed an action a short moment before it reached the end state. Most of the photos were totally new to Alex. 3) Live condition: Actions (10 non-repetitive and 5 repetitive actions) were demonstrated live.

The important finding of this experiment, apart from the fact that similar results were obtained in condition 1 and 3, is that Alex identified and performed the actions equally well in conditions 1 and 2, irrespective of whether the actions were repetitive or not.

A semiotically "rich" interpretation of this result could be that Alex not only used the picture as a sign for the real-world action, but that he could simultaneously recognize a complete action including its goal state from an earlier phase of its development, i.e. that he was capable to grasp a form of indexicality (in this case *temporal* contiguity) indirectly through the sign. However, the first factor which argues against such an interpretation is the lack of any evidence concerning novel actions. Since all actions involved were taken from the set of actions on which Alex had been trained earlier on, and since we also are familiar with the difficulties Alex had been having with the imitation of novel actions, we cannot exclude that a much more simple explanation in terms of *stimulus enhancement* could be given. This is aggravated by a second observation: what has here been termed means and ends are really part of what, in the Lifeworld, is a continuous stretch of behaviour, which means that the two phases (the final and the pre-final one) are not clearly differentiated. It may seem that, if Alex sees the picture *as* a picture, that is, not according to reality mode, the content of the picture must be differentiated from that phase of behaviour which he goes on to execute. However, it is perhaps not impossible for Alex to differentiate expression and content, without doing the same in the relation between the piece of action depicted and the overall action which he himself realises. Still, even though the performance can be explained most parsimoniously in terms of stimulus enhancement, the successful recognition of the incomplete representations added support for the interpretation that Alex did indeed regard the stimuli as "pictures" rather than as simple cues (cf. 4.3).

4.2.5 Conclusions

Due to the relatively few experiments and subjects involved, the studies related to ape capacities in imitation carried out within SEDSU cannot answer conclusively the question stated at the onset: which species are capable of it, and to what degree. However, they strongly suggest that even precursors of imitation such a yawning contagion cannot be taken easily as "automatic responses", and that action imitation is difficult – even after extensive training and for a (partially enculturated) chimpanzee, and that this is not due to lack of

motivation. While comparisons with human subjects, and between apes species for all tasks, were not performed, based on the results and findings reported in the literature, we can order the capacities attested in an *evolutionary sequence*: (a) recognition of others' imitation, (b) yawning contagion, (c) imitation of object-related actions (d) imitation of novel bodily movements. Chimpanzees are capable of (b), enculturated ones of (c), but only human beings have been shown to be proficient in (d).

4.3 Non-conventional signs

The concept of non-conventional signs remains somewhat ill-defined - even though, taking into account our definition of the sign, it has been a lot more specified here than ever before, for example in relation to classical notions such as non-verbal communication, non-verbal signs, etc. There is of course no way for us to cover the immense domain of different kinds of non-verbal signs (even excluding meanings which are not signs). However, the type of non-conventional signs on which SEDSU research focused most was *the picture sign*. Other kinds of non-conventional signs, especially *gestures*, will be discussed in Section 5, since, in the SEDSU project, these studies focused on the development of the relevant capacities in children.

4.3.1 The picture sign

A picture is a surface equipped with markings giving rise to a vicarious perception of objects and actions of the perceptual world (Gibson, 1982). In order to see a picture *as* a picture, i.e., *as* a sign, it is necessary to perceive at the same time the similarity and the difference between the surface and that which it depicts; this, according to Gibson, is a capacity only found in human beings. However, this needs to be empirically investigated rather than being taken for granted. The picture is a sign, which is of the general type of primary iconical signs. It is therefore different from droodles (configurations of lines which may be identified as representing some object or fact once they have been said to represent them, that is once a "key" has been given). What pictures represent is always percepts, that is, scenes from the real perceptual world. The content of the picture sign is therefore often said to be quasi-perceived (by e.g. Husserl, Gibson, Wollheim). The expression of the pictorial sign is a plane surface, but what is "seen into" it is a three-dimensional scene of the perceptual world. Primary iconical signs such as pictures seem to presuppose a distinction between two-dimensionality and three-dimensionality (Sonesson, 2000), which has independently been shown to be difficult to grasp for at least some non-human primates (Barbet & Fagot, 2002). Pictures are also particular in allowing for a one-to-one mapping (to a limit) of single traits of their expression to single properties of the perception, which they render. But this one-to-one mapping is only possible through the intermediary of an identification of the whole (termed "resemanticisation" in Sonesson 1989). It is only because a particular circular figure is seen to represent a head, that a line that stands out can be perceived as a nose, and so on.

It will be noted that scale-models, dolls, dummies, people impersonating other people, and the like, are not pictures in this sense. It has been shown by Deloache that scale-models are more difficult for children to understand than pictures: not only is the picture understood later than language in these experiments, around 2 1/2 years (Deloache & Burns 1994, etc.), but scale models are understood even later, at 3 years of age, half a year after pictures (DeLoache 2000). This result was predicted by Sonesson (1989), who argued that the discovery of the similarity between one three-dimensional object and another was contingent of the realisation of their sign character. Sonesson also suggested (something which remains to be shown) that it should be even more difficult to discover the sign character of animate beings. Thus, theatre requires an elaborate system of conventions.

More recent experiments have shown that even children 5 months of age look longer at a doll than at its picture (DeLoache & Burns 1994). However, it does not follow from this that the children see the picture as a picture. Indeed, 9 months olds, but not 18 month olds, try to grasp the object depicted as if it were a real object (DeLoache 2004); whatever the difference they perceive, then, it does not seem to involve signs as opposed to objects. This result shows that the picture and its object are seen as being different, but not necessarily as forming a sign-vehicle and its referent. The real doll is perhaps seen as a more prototypical instance of the category; or, alternatively, the real object may be more interesting because of having more perceptual predicates.

In all of Deloache's experiments, the task is, in one way or other, to find a hidden object by using information contained in a picture or a scale model. The experimenter and the child are at first outside the room in which the child is to search for the toy. The child cannot see the picture or scale model and the room at the same time. The experimenter tells the child that she will hide the toy in the room and then come back and ask the child to retrieve it. She returns to the child and points out the appropriate location in the picture/scale model telling it "This is where Snoopy is hiding in his room. Can you find him?". If the child fails in the first search he or she is once more shown the picture and given more explicit prompts. 24-month-old children do not pass the retrieval test, but 30-month-olds do; there is no difference in performance using photographs or line drawings. However, when the whole procedure is conducted verbally, children pass the test already before 24 months; and when a scale model is used, only 36-month-olds pass it.

This way of investigating the picture function may be criticised from two diametrically opposed points of view. First, it could be argued that the task involves much more than the recognition of the picture as picture – it requires a particular action that is no doubt difficult in itself, namely, to search for the hidden object.[9] It remains, however, that even this task is differently accomplished if the instructions are given entirely in verbal form, or if they involve pictures or scale models. In fact, it turns out that children who are unable to retrieve the object in the room can still identify it from one picture to another (Lenninger, in preparation). Such an identification appears to involve the identification not of areas on surfaces, nor of real things, but of pictorial contents (two depiction of the same doll). What appears to be difficult, therefore, at this stage is not the step from the expression to the content, but the further step from the sign to the referent. Language earlier on than pictures seems to be directly connected to the real world.[10]

On the other hand, even when the instruction for the task features pictures or scale models, at lot of verbal and indexical scaffolding also takes place, without this being taken into account in the interpretation. It has been argued by Callaghan & Rankin (2002) that pictures would be interpreted even later if such verbal scaffolding did not occur. More fundamental, however, may very well be the indexical scaffolding: not only are the objects *pointed out* by the experimenter in the picture or the scale model, but the latter are even placed *on* the real objects, creating an artificial neighbourhood relation. Indexical scaffolding turned out to be an even more important factor in the study of picture understanding by apes (Persson 2008).

Another of Deloache's experiments seems to indicate that understanding the sign function is at least part of the problem. When the experimenter, instead of talking about a model and a real room, tells the children that the search has to take place in the same room, which has shrunken since it was last seen, the task is accomplished much more easily (DeLoache & al. 1997). The difference, clearly, is that the two instances are here connected by a narrative chain rather than by a sign relationship. In another experiment, DeLoache (2000) places the scale model behind a window-pane, in order to make it more similar to a picture, with the expected results. In fact, however, two things happen here which would have to be separated: the object becomes less prominent, because it has less the appearance of three-dimensionality; and it is put into a frame, which creates a centre of attention.

If understanding pictures is as difficult for children as DeLoache and, even more, Callaghan, suggest, then we should hardly expect animals to be able to do so. Most

[9] We owe this suggestion to Sara Lenninger, who is preparing a dissertation on this subject at the Department of Semiotics in Lund University.

[10] In a SEDSU experiment (D13, experiment 2), capuchins were shown to be better at picture-to-picture matching than matching pictures to objects – although they were better at going from objects to pictures than vice-versa. However, this could easily be a case of surface mode or reality mode – though the fact that the direction from object to picture different from the opposite direction may indicate otherwise. This was not confirmed in a later experiment (D13, experiment 3).

experiments purporting to demonstrate the ability of some animal species to interpret pictures have neglected to investigate whether the animals are also able to tell the difference between the picture and its object; but Cabe (1980:335), who makes this observation, states he has taken pains to ascertain that the pigeons of his experiments possess the later capacity (p 313f). Yet it is of course possible, again, that this mixture of the same and different behaviour in relation to expression and content simply reflects a prototypical ranking within the category of real things, or a difference in perceptual richness. On the other hand, some primatologists tend to take it for granted that apes involved in language-training studies already understand pictures (cf. Persson 2008). There are only a couple of investigations specifically dedicated to apes looking at pictures and scale models. Itakura (1994) reports that enculturated chimpanzees can interpret line-drawings. Moreover, Kuhlmeier & al (1999: 2001; 2002) have even shown that their chimpanzees, going through the same kind of experiment as that presented to children by DeLoache, are able to understand scale models. It is difficult to know what to make of these results, since the apes involved in these studies were more or less "enculturated", which is to say that they have been made familiar with many of the semiotic resources that in ordinary circumstances are peculiar to the human lifeworld (e.g. Kanzi is reported enjoying TV-cartoons). Moreover, it should be noted that, while the children in DeLoache's experiment were introduced to the model of a room that they had never seen before the training-phase, the apes were confronted with a model of their own familiar environment. In addition, a lot of facts about the subjects and the experimental procedure are not clear from the published reports.[11]

4.3.2 Pictorial competence in non-human primates

Several experiments realised within the framework of the SEDSU project have some bearing on the study of the capacity of different primates for grasping the sign function of pictures (D13; D17; D21; D22; D24). In spite of the Itakura and Kuhlmeier studies mentioned above, the general impression so far has been that picture interpretation, when really investigated as such, is very difficult for monkeys and apes (Persson 2008) – as well as for small children, according to the studies of Deloache and Callaghan. From the point of view of the SEDSU project, it is important to find out *why* picture understanding is so difficult for even our closest kin on the evolutionary family tree, and also for small children. One hypothesis could be that perceiving at the same time the similarity and the difference between the picture and its content/referent may be very difficult for nonhuman subjects (such as non-human primates) with no language abilities as well as for human subjects (such as *some* children with autism and congenital deafness) with limited or delayed linguistic abilities. An alternative hypothesis is that picture understanding does not depend on language per se, as long as subjects have attained the sign function. Nevertheless, people with certain impairments may process pictures differently than controls. In this sub-section, we discuss the studies of pictorial competence involving non-human primates, while in the next sub-section we turn to the impaired populations.[12]

Experiment 1 (D13) is particularly relevant and has three main original features. First, it assessed picture processing in nonhuman primates by way of an experimental design aimed at testing picture/object confusion. Second, it used pictorially naïve primates as subjects. Third, subjects were selected from three species strategically selected within the primate phylum: 55 anubis baboons (Papio anubis) from the CNRS Rousset-sur-Arc Primate Center (France), 7 chimpanzees (Pan troglodytes) from the Wolfgang Kohler Primate Research Center of the Leipzig Zoo, and 4 gorillas (Gorilla gorilla) from the zoos of either Leipzig or Nuremberg

[11] More about picture perception in pigeons, and some in apes, can be gathered from Fagot, ed. 2000. However, these studies do not take into account the difference between the picture and the depicted object.

[12] The experiments involving biological motion by children with autism and deafness (D17), as well baboons (D13 and D21) were discussed in Section 4.1, since we consider them to show interesting results concerning visual perception, but that do not bear on the question of pictorial understanding.

(Germany). The stimuli consisted in pebbles or slices of banana, which were presented either real or in photographic pictures. Pictorial representations of the pebbles and banana slices were in colour and subtended the same size as the real objects. They were displayed on white 8 cm X 8 cm backgrounds. Pictures were printed on a matte photographic paper preventing surface reflectance. A two-alternative forced choice procedure was used. In addition to noting the choices, four behaviour types were coded: grasping/touching, smelling, bringing to mouth, and eating.

Among the 66 subjects (all species confounded) 47 chose the real banana and only 19 chose the pictures of the banana when the real banana was opposed to its picture. Considering the alternative of the banana picture and the real pebble, the analysis of stimulus choices for the remaining 55 primates who reached for one stimulus indicated a reliable bias for the picture of the banana slice: 44 primates preferring the banana picture, and 11 preferring the pebble. However, this reliable preference was not replicated in the chimpanzees. The participants also revealed strong preference for the picture of the banana over the picture of the pebble. Again, chimpanzees as a group showed no reliable preference for either the banana picture or for the pebble picture.

The behaviour type "eating" was one of the most frequently observed ones. It was observed in 39.5% (i.e., 17 out of 43) of the baboons who chose the picture of the banana, and 100% of the gorillas (i.e., 3 of 3). On the other hand, eating was never observed in chimpanzees, in spite of a larger number of chimpanzees than gorillas, therefore suggesting species differences in spontaneous responses. These results were confirmed in a second bout of manipulation. (Cf. Figure 4.6).

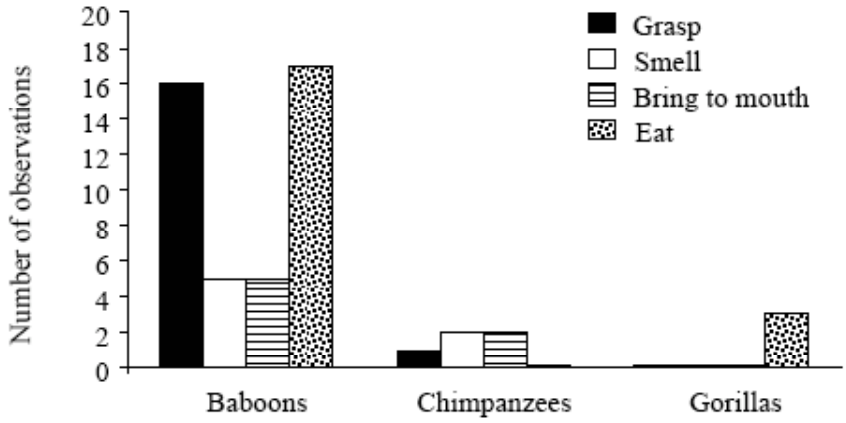


Figure 4.6. Observed frequencies of each action during the first bout of manipulation

In contrast, the picture of the pebble was only rarely chosen, and none of the primates reached it twice during the experiment. In addition, the analysis of the behaviours elicited by the pictures of the pebble revealed that this stimulus was almost *never eaten*, contrary to the banana pictures. Of the 9 baboons who took the picture of the pebble 4 manifested touching, 3 smelling and 2 bringing to mouth. This finding should be contrasted with the large number of baboons who ate the picture of the banana slice (i.e., 17/43 for the first sequence of manipulation). None of the gorillas took the picture of the pebble in these trials, confirming that gorillas expressed much less interest for the pictures of the pebble than for the picture of the banana slice. Three chimpanzees took the picture of the pebble once, one smelling it, one bringing it to mouth, and the third eating it. In sum, 35 eating behaviours were observed during this experiment and all of them, except one, were observed when the subjects manipulated a picture of the banana. All species combined, the picture of the banana slice was

therefore eaten reliably more often than that of the pebble.

Given these differences between the real piece of banana and their photographs, we may interpret this as indicating that the baboons processed the pictorial pieces of banana as poor exemplars of the category, and preferred the real food over its depiction for being closer to being a prototypical banana. Alternatively, the real banana was more interesting for having more perceptual properties confirming its membership within the category of bananas. In the next test trials, poor typicality did not prevent preference for the picture of the banana to prevail when it was contrasted with a pictorial pebble. Gorillas behaved similarly to baboons (This could however be an artefact of the small number of gorillas tested). Chimpanzees behaved very differently from baboons and gorillas in our task, and the data set provides no evidence for confusion in these animals. Thus, these animals expressed no reliable preference at the group level for either stimulus in the tests, suggesting that the banana and pebble pictures were similarly attractive to them. In addition, and contrary to gorillas, no single chimpanzee could be observed eating the picture of the banana during our task, although more chimpanzees were studied than gorillas. These two results suggest that the chimpanzees did not confuse the real piece of banana and its picture. Of course, this conclusion does not necessarily imply that chimpanzees processed our pictures as signs. Yet it could be taken as suggestive of a tendency for sign use, when compared with the behaviour of the other monkeys and apes.

The results of some other experiments could be taken to suggest that even monkeys are able to perceive pictures as pictures under appropriate circumstances. However, in these cases other interpretations seem more plausible. In one experiment (D13, experiment 5), capuchins were tested on the matching of stimuli consisting of two pairs of real objects: a cork and a stylized wooden apple; a clothespin and a cigar-lighter, as well as the corresponding black-and-white photographs, silhouettes and line drawings of the same objects (see Figure 4.6). Several analyses were made comparing the conditions, which all gave similar results. Thus, for instance, an ANOVA test comparing the overall mean percentage of correct responses obtained in the three picture-picture control conditions and in the three object-picture testing conditions showed that performance levels did not differ across conditions (photograph-object, 90.6%; line drawing-object, 85.9%; silhouette-object, 86.4%; photograph-photograph, 95.3%; line drawing-line drawing, 92.2%; silhouette-silhouette, 90.6%). All in all, results demonstrate that transfer performance remained high, whatever the condition.

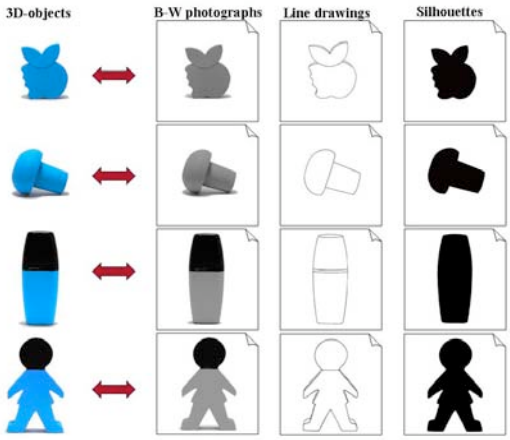


Figure 4.7. Test stimuli in the transfer tests of experiment 5 (D13)

Results like these could be taken to mean that the animals, with appropriate training, are able to associate objects with their depictions, even if degraded depictions are used, such as silhouettes or drawings. If one goes on to claim that line drawings do not exist in nature, and

that they are very different from the real objects they represent, one is naturally brought to the conclusion these animals did not confuse the drawings with their real referent. If so, it would have been demonstrated that the capuchins did treat the drawings as iconic representations of the real objects. However, as has been convincingly argued by James Gibson (1966; 1978; 1980; 1982), while lines may not exist in nature, there do exist higher-order variables which are common to lines and to the limits between different areas of light offered to the eyes, in particular when there are two of them and they are in constant movement. Thus, this could very well be a case of holistic (probably global) perception, as discussed in Section 4.1. This being so, the experiment cannot demonstrate the kind of asymmetric relationship between expression and content characteristic of the sign function. Again, we could very well be faced with different exemplars (perhaps more or less prototypical, but not necessarily so) of the same category.

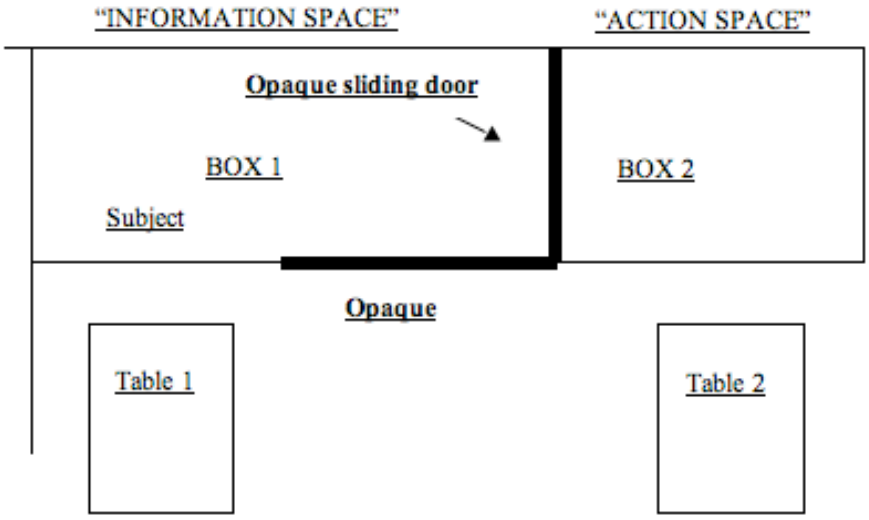


Figure 4.8. Experimental set up of the video experiment

Another SEDSU study (D13, Experiment 9-10) was designed to assess whether capuchin monkeys were able to use video clips to gain information about an event in which objects were hidden. During training, subjects were in BOX 1 (Figure 4.8), and could see the experimenter "hiding" food under one of two objects on Table 1. The subject could retrieve the food items only after moving to another similar test chamber (BOX 2) and choosing between two objects on Table 2, which were replicas of the objects shown on Table 1. There were two conditions: a No-shift Condition, in which the position of the two objects with respect to the subject in the searching phase and in the hiding phase were the same, and a Shift Condition, in which the position of the objects in the two phases was reversed. The saliency of the objects was increased by painting a white frame enhancing their specific shape. Subjects' performance was initially at a random level, but eventually most learned to search under the "right" object.

During testing, at the beginning of a trial the subject was similarly in BOX 1 and looked at a monitor placed on Table 1 as a pre-recorded video-clip was shown of the experimenter hiding food under one of two objects. Then, the monitor was turned off, the opaque sliding door was opened and the subject moved to the next adjacent BOX 2 to choose between two real objects placed on Table 2. These objects were identical to those shown on the video clip. According to the hypothesis of the experimenters, if the subjects understood the relation between the virtual objects shown in the video and the real objects that they had to select, the subjects would search the domain at above chance level. If, on the contrary, the subjects did

not understand the sign-referent relation, the subjects would have no way to know where to search and they would search at chance level. The results showed that after 16 sessions, one subject not only learned to use video clips appropriately, but also generalized to new objects. This finding may be taken to show that the ability to use video clips to gain information about real objects is within the capacity of capuchins.

However, this finding does not demonstrate that the video was perceived by the capuchin as a representation (sign). In fact, the capuchins may well have taken the video monitor to be a window permitting a view into another room. Indeed, this may even be the most natural interpretation, given the placement of the video screen and the way the two rooms had been connected. As has been argued by Persson (2008), commenting on other, somewhat similar experimental designs, learning to understand "the rules of the game called window" is no doubt primary to understanding such more difficult games as interpreting pictures and mirrors (Cf. Sonesson 2003b). In the case of the mirror, the object is really where is appears to be; it is only not reachable by taking the way straight ahead, but instead it is necessary to use a way around the almost invisible obstacle. Both ordinary pictures and mirrors are different, in somewhat different ways, in placing things visually where they do not appear for the other senses, in addition to blocking the direct way to investigation. Indeed, the capuchins, instead of learning to use pictures, learnt to go into the other room to reach the objects at the other side of the window, which is comparable, *mutatis mutandis*, to the children understanding the shrinking room in the DeLoache experiment, but not the sign relationship. While it is not possible to know what exactly the capuchins where "thinking", it is sufficient to suppose that they went in the direction in which the table they had seen earlier could be reached, but happened upon another table, which, in all relevant respects was identical, before arriving there. Just like the children were not surprised that the room could shrink, the capuchins apparently had no difficulty accepting that the table was in another place.

In a variation of this study, scale models were used (D26). Just as in the case of the video studies, food was hidden under one of two different objects at each trial. The objects functioned as landmarks directly marking the food site in turn across trials. When capuchins observed directly the hiding events they had no problem at all to retrieve food. Capuchins could also successfully use information about the hiding event obtained from non-real sources such as scale models of landmarks or videos of the hiding event. In fact, only one capuchin solved the video tasks. Once again, it was not clear that anything like picture understanding is involve in these studies. What was described as the use of scale model is really the situation in which, after the objects have been hidden, the same space is again made available to observation by the monkeys, but the two objects observed by them before have now been exchanged for two identical, but much bigger objects (left and right position is varied, so that this cannot be the determining factor). It is not clear that any sign relationship comes to the fore in this study. Rather, we are faced with two expanding objects, even more obviously parallel to the "shrinking room" in DeLoache's studies, by a simple invertion of the size relationship: "an expanding room".

The set up of many of these experiments would seem to take for granted that all primates can learn to use pictures, given a sufficiently extended and progressive training to associate the real objects and their depictions. That such a training, in the case of human beings, is really implicit in Western culture has often been taken to be proved by cross-cultural studies showing that humans from remote cultures with little to no experience with pictorial displays are often unable to recognize the objects depicted on photograph (Deregowski, 2000). In fact, however, it was shown already by Kennedy (1974:69ff) and Jones & Hagen (1980:203 ff) that such cross-cultural evidence is actually non-existent: what is taken as such is either unconfirmed anecdotes counted by travellers, or studies made in South Africa during the apartheid period, without taking into account the social consequences of having a white experimenter posing questions to native people, although these are evident

from the fact that some persons waited an hour before making their reply. Ironically, Deregowski himself come up with the explanation for one of the aforementioned anecdotes, the study by Muldrow concerning the Me' tribe, whose members smell and taste the pictures, but do not think of looking at them. According to Deregowski (1973:167; 1976:20) not only pictures, but materials like paper are unknown to the Me'; therefore, when Deregowski had pictures printed on coarse cloth, animals well-known to the tribe could be identified, although the recognition was still not immediate. In the case recounted by Muldrow, it seems the Me' were so busy trying to discover the fundamental properties of the paper as an object in itself, that the iconic properties, those making it a pictorial sign of something else, were not noted; other attributes became *dominant* in their experience of it. It therefore seems (as suggested by Sonesson 1989) that for something to be a pictorial sign of something else, it must occupy some relatively low position in the particular Lifeworld hierarchy of "things". If so, whatever the training, human beings, may well possess an ability for picture interpretation which will never emerge in some species, such as monkeys. Indeed, this has been confirmed in some studies of picture interpretation in apes, where the salience of surface properties turn out to be the important factor (Persson 2008).

The results of most pertinence for the study of pictorial competence in a non-human, primate, however, emerged from a number of studies that were concerned with imitation, and which involved the chimpanzee Alex, who was trained and tested within the Do-as-I-do paradigm (see Section 4.2.4). 20 of the training actions that Alex had mastered in the previous study were selected and Alex was solicited to perform them, not by a live model as before, but by being shown the actions on video, a colour photograph, a black and white photograph and a drawing depicting the target actions (D24).

The results showed significant differences between all conditions, but Alex performed above chance in all of them, except on the drawings. The experiment was repeated with different sized pictures, as well as with black-and-white as opposed to colour photographs, without finding any differences. Finally, the task was conducted with pictures representing still actions with an incomplete goal (incomplete) as well as pictures of the same action in which the goal had been achieved (end state), once again without any significant difference between the two pictorial stimuli, while these had fewer correct responses compared to a live model.

In the cases when the action was shown on video, it is not possible to say whether the live illustration of the action and the video were qualitatively different for Alex (although there was a quantitative difference in the first experiment). However, a still photo serving as a prompt for a real action must certainly be considered different from the action, at the same time that it appears to have been taken by Alex to "stand for" it, as shown by the fact that he performed the represented action. To suggest that Alex is simply confusing the still photo, and even more the photo of the incomplete action, where the picture prompting the action is two times away from the action requested, seems indeed quite improbable. It is therefore possible to conclude that picture understanding is within the purview of chimpanzee capacities, and since Alex was neither language-trained (nor engaged in any other form of sign use), that *it is possible to understand pictures as primary iconic signs, quite independently of language*. However, we should note that Alex had extensive experience with imitation both prior to the study, and during these studies (see 4.2.4). This is consistent with the close conceptual link between imitation and sign use, argued for in Section 2.

Finally, it is not completely clear why Alex would reproduce actions depicted in complete or incomplete photos but not actions depicted in drawings. One possibility is that he did not see the drawings as representation of actions that he was required to reproduce but merely as a series of lines on a white background, i.e. he operated in the a-mode of Daddesio (1995) or the "surface mode" of Persson (2008) (see Section 2.2), due to the degraded nature of the representation.

4.3.3 Pictorial competence in children with impairments

A number of studies investigated the way children and adults with impairments known to affect linguistic competence, such as ASD and deafness (without sign language), processed and interpreted pictorial stimuli. The questions were (a) whether at least some of the subjects may have difficulty to process these in pictorial mode (b) whether such an impairment could be explained as resulting from delayed or impaired language and (c) whether there were other peculiarities in their processing of (certain kinds of) pictures, in relation to their known perceptual difficulties, especially in relation to social-emotional aspects (cf. Section 2.1)

In one study Da Fonseca et al (D17, Exp 4), compared the interpretation of cues for emotion vs. the identity of objects in adults and children with autism, in relation to typically developing controls. 60 colored photographs taken from popular French media magazines were chosen, scanned and used as visual scenes in which either a face expressing an emotion or an object was masked. Visual scenes masking an object or face were carefully matched in terms of complexity and the number of characters and objects contained in the scene. Target faces and target objects were masked by a white circle (about 2 cm of diameter). Participants were asked to infer the emotion of the hidden face or the nature of the missing object, and in both cases, three choices were given (in the case of emotions: three "emoticons"), only one of which was correct. The experimental procedure presumed that participants could understand the representational nature of pictures and emotions, and indeed, they were soon at ease with the procedure.

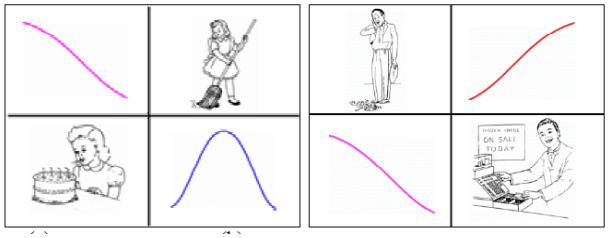
The results showed that children with ASD were poorer than typically developing children in the Emotion condition, but not in the Object condition. This deficit is difficult to interpret in terms of an overall inability for relying on contextual cues, as postulated by "weak central coherence" theory, because it should then have affected both the Emotion and Object conditions. Furthermore, the adults with ASD performed on the same level as controls for both conditions. These findings show not only that abilities to interpret pictures may expand with age in ASD participants, but that the relevant abilities were not a matter of direct perception – which does not involve inference - but rather a manner of inferential interpretation consisting in the "filling in" of the missing element based on relevant world knowledge. Indirectly, this further supports the interpretation that the stimuli were being processed in "pictorial mode".

Roset et al (D17, Exp 5-6), compared the recognition of emotion in three kinds of representations of faces (a) photographs of human faces, (b) cartoons of human faces and (c) cartoons of non-human phases, in two conditions: upright and inversed, in children with autism, deafness and typical controls. As is well-known, inversed presentation of faces requires higher degree of local processing, and a different rate of correct interpretation in the upright vs. inverted conditions would imply that configural processing was used in the ordinary, upright view of the faces. Typical and deaf children both showed a clear negative effect from inversion, for all types of face representation. Interestingly, the autism group also showed a similar effect, but only for (b) and (c), i.e. the cartoon faces but not the photographs. This implied that they used a configural strategy for the cartoon faces, but a local one for the photographs. The relevance of these findings are the following. First, children with ADS are not generally impaired in configural processing, but mostly in the cases where this involves realistic social-emotional contexts. Secondly, the fact that schematic pictorial representations such as cartoons may be processed similarly to controls (but differently from their own processing of photographs of human faces), implies that there is no general impairment in pictorial competence in people with autism.

A complementary study using a similar design was performed by O´Konnor et al (D17: 7-8). Here again, cartoon vs. photograph representations were compared in upright vs. inverted positions, but the task was not one typically requiring configural processing such as emotion-recognition, but one thought to stimulate local processing: discriminating between

two nearly identical pictures on the basis of minor differences (in the eyes or mouth of the face). Typically developing children were compared with children with ASD, as well as a group of typical adults. The results showed that typical adults found it easier to discriminate the photographs in the upright condition (when a configural strategy could be applied), but the cartoons in the inverted condition (which stimulated a local strategy). The comparison between the two groups of children showed greater accuracy for photographs than for cartoons in the typically developing, but not in ASD participants. This finding confirmed that cartoon faces are processed differently from photographs in typicals, but unlike the previous one involving emotion recognition, suggested that for the purpose of feature-based discrimination, ASD children processed both types of faces similarly (apparently locally).

In a final study here summarized, Järvinen-Paseley and Heaton (in press) presented to typically developing and autistic children 24 utterances pronounced with four distinct pitch contours. Twenty-four visual response slides were then constructed, with each including (1) the correct pitch contour, represented iconically as a curve the shape of which matched the pitch of the utterance, (2) an incorrect pitch contour representation, (3) a picture which matched the semantic content of the utterance, and a (4) a picture which did not match the semantic content (Figure 4.9). The positioning of correct perceptual and semantic targets was randomized across slides. During presentation each utterance was followed by a visual display.



(a) (b)

Figure 4.9. Examples of visual slides used in the experimental stimuli, for sentences (a) "I like growing older", and (b) "I will lose my job"

Training preceded the administration of the experimental task to ensure that all children could perform the auditory to visual mappings required. For the perceptual training, the child was told that sentences could be said in different ways, so as to form differently sounding shapes, depending on how "high" the voice sounded. The child was then shown the visual display on the laptop computer, and told that his or her task would be to point to the shape that s/he thought best matched each sound. For the semantic training block, the child was further informed that each of the sentences told a little story that could be depicted in a picture, and that the experimenter was going to play the sentences again, but this time the child's task would be to point to a picture that s/he thought best matched the meaning of each sentence.

The results showed that while both autistic and typical children provided more semantic than perceptual (prosodic) interpretations, the autism group provided significantly more perceptual choices that the controls. Furthermore, they did so more accurately than the controls. The children with autism were as accurate in their semantic as in their perceptual choices, while the controls showed significantly poorer accuracy in their perceptual than semantic responses. The implications of this are (a) that children with autism have no difficulty interpreting even fairly abstract pictorial stimuli such as pitch contour representations, and (b) that they can perform a cross-modal mapping in such cases even more accurately that typical children.

The following main conclusions can be drawn from the series of experiments summarized in this sub-section: (1) Autistic participants (children or adults) are perfectly

capable of properly interpreting pictorial signs. (2) They can interpret pictures in a sophisticated manner, and for instance infer what is the missing part of a complex visual scene and match the perceptual attributes of linguistic signs to pictorial-abstract representations of theses perceptual attributes. (3) When a deficit emerged in autistic children or adults, their deficit was systematically restricted to the processing of the socio-emotional content of the pictures – but only when these involved photographs and not cartoons, which seemed to be processed in a configural rather then a local mode. (5) Deaf children demonstrated a normal ability to understand pictures, and are not particularly impaired in that respect. (6) The initial rationale for including deaf and autistic children in these studies was their common known deficits in both social communication and language acquisition compared to controls. The lack of any general problems involving picture processing in these two populations, suggests that this ability is not so closely linked to intact language.

4.3.4 Conclusions

The main conclusion resulting from the research described in this section is no doubt the confirmation of the idea that at least higher apes may be able, with some difficulty, to understand pictures as pictures; this ability is, so to speak, in their zone of proximal evolution. However, even when apes are seen to prefer the picture of a banana to a real pebble but a banana to its picture, this may perhaps be more parsimoniously explained by a perceived difference between prototypical bananas and things that, for lack of better exemplars, may be considered as some kind of banana at the limit of the category. In the ordinary world of apes and monkeys, markings on surfaces are probably not sufficiently salient in themselves, as Gibson suggested long ago. The studies of children with autism and deafness showed the kind of deficit which would have been expected from what we know about the nature of their problems: none in the case of deaf children, and difficulties involving social and emotion-laden content in the autistic children. This suggests that language use is in no way critical for the interpretation of pictures. Given this result, it is a pity that we could not, within the SEDSU project, study picture comprehension in normally-developing children, in particular since the studies which have been done so far are very much dependant on a linguistic paradigm. It seems quite possible that the late onset age of picture understanding in the studies by DeLoache is at least partially due to the kind of task used: the retrieval of a hidden object. To accomplish this act, it may not be sufficient to see the picture as picture, but one must also understand that it is about something in the real world, much like in language.

4.4 Conventions – their precursors and effects

The definition of convention adopted in Section 2 was "*(a) a regularity in behaviour, (b) that is commonly known in a community and (c) has an irreducibly normative nature* (i.e. members of the community expect other members to abide by it)." A symbol is a conventional sign, i.e. one where the connection between expression and content is commonly known, e.g. the signs on toilets representing in various ways that certain toilets are for men and others for women (or both). The fact that these signs also have an iconic ground does not make them non-symbols, but simultaneously signs of secondary iconicity (cf. Section 2). The same can be said of most of the signs of language: while many have both iconic and indexical aspects, if it was not for their conventionality, language would not be able to serve as a communicative medium.

Conventionality and symbolicity must therefore be distinguished from both *arbitrariness* (at least as popularly understood in the sense of "unmotivated" and "anything goes") and *ontogenetic ritualization* (e.g. an infant's habit of lifting up his arms when wanting to be picked up, cf. Tomasello 1999). On the other hand, the two are closely related to intersubjectivity (Section 2), since the latter implies experiential sharing, and the ability to entertain common knowledge constitutes an advanced form of such sharing. In the developmental framework of Bråten (2007), *primary intersubjectivity* (more or less from birth) involves mutual gaze and shared affect, *secondary intersubjectivity* (from appr. 9 months) is about objects in the environment and *tertiary intersubjectivity* (from appr. 24 months) is based on "symbolic conversation with actual of virtual others" (ibid: 3). In other words, the most advanced form is based on the ability to share conventional meanings and to use them in "self-other dialoguing in dramas of narrative imagination, for simulation of conversational partners' minds" (Bråten and Trevarthen 2007: 23).

In this section, we summarize *7 comparative studies* concerning intersubjectivity and conventionality. The first attempted to test symbolic capacity in capuchins, but used a definition of symbols based on the "arbitrary" connection between token and food items (D13; Addessi et al 2007). The second compared different species of great apes with respect to parental "tolerance" to their infants (D19). The third and fourth compared intersubjective behaviours (and indirectly capacity) in ape and human infants with respect to primary (D19) and secondary intersubjectivity (D7; Zlatev et al 2008).

The last three were cross-cultural and cross-linguistic: the fifth investigated the role of different conventions (norms) on pointing for children in Sweden and Thailand (D19; Zlatev et al in press), the sixth asked if different conventional ways of expressing motion in French and Swedish resulted different spontaneous categorizations of motion situations and the last explored the role of cultural and linguistic convention for the conceptualization of time in Amondawa, in comparison to the West.

4.4.1. Relative quantity judgments by capuchins

A series of experiments investigated the use of valueless objects such as coloured chips as possible "symbols" by capuchin monkeys (D13, Experiments 16-19). The paradigm used was the exchange of tokens, i.e. inherently non-valuable objects that *acquire an associative value* upon exchange with the experimenter (Brosnan & de Waal 2004). The goal was to assess to what extent capuchin monkeys deal with tokens as they do with real food. 10 capuchins firstly learned the association between a given token and the corresponding amount of reward returned by the experimenter when the token was exchanged. Each subject learned to exchange two different inherently non-valuable objects of similar dimensions (but differing in shape, material, and colour): token A and token B. Token B was exchanged for three rewards, and token A was exchanged for one reward (a reward consisted of 1/8 of a peanut seed).

In a first experiment it was assessed whether capuchins chose the larger between two

quantities of the same food. Each subject was presented with a "choice apparatus", constituted by a platform with two sliding trays where different quantities of peanuts (one offer on each side) were available. The subject could choose one of the two offers (by pulling one of the two sliding trays) on the basis of the amount of food presented. In each 20-trial session, all the possible pair-wise choices (N = 10) between 1 and 5 pieces of food were offered. Criterion of 90% correct performance was reached in 5 sessions on average (range 2-11). Performance decreased with decreased difference between the quantities, but not with overall quantity. A second experiment using tokens instead of food items showed similar results, though with a longer number of sessions necessary to reach criterion.

The crucial experiments tested whether the two different associations (A – 1 reward, B – 3 rewards) were stable enough to allow a "maximization of payoff". For each condition (1B vs. 1A, 1 B vs. 2A, 1 B vs. 3A, 1 B vs. 4A, 1B vs. 5A), it was evaluated whether the percentage of choices for the token B differed significantly from a theoretical chance distribution with a mean of 50% by means of a one-sample t-test. Moreover, to evaluate capuchins' initial performance, Addessi et al. analyzed the frequency of correct choices (i.e. token B in the conditions 1B vs. 1A and 1B vs. 2A, and tokens A in the conditions 1B vs. 4A and 1B vs. 5A) in the first 12 trials (corresponding to the first 3 sessions) of each condition by a binomial test. All analyses were performed at the individual level, since subjects showed different choice strategies. As shown in Figure 4.8, the 10 subjects behaved according to *three main strategies*: (a) Four capuchins maximized their choice by selecting the token B when this was advantageous, i.e. when one token B was presented against one or two token(s) A, and selecting tokens A when this led to more than the three rewards of token B (i.e. when 4 or 5 tokens A were presented against one token B); (b) Other four capuchins consistently preferred token B regardless of condition. (c) Two capuchins significantly preferred token B only when presented against one token A; in all other conditions, they significantly preferred tokens A.

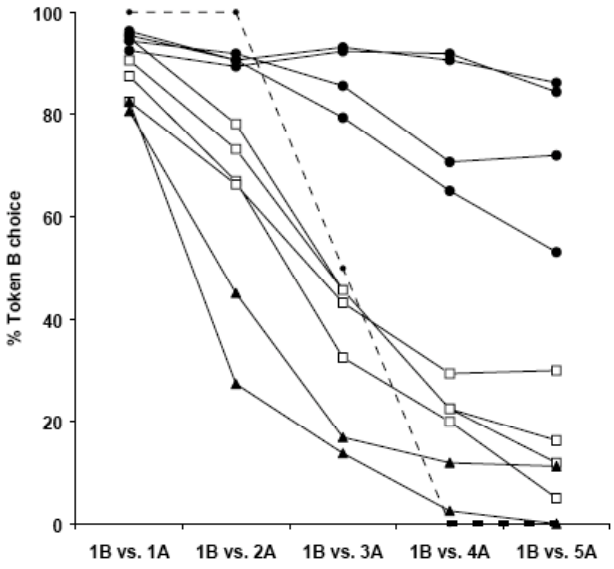


Figure 4.10. Percentage of token B choices for each comparison and subject. The dotted line indicated the ideal curve. Strategies are indicated as follows: only B (black circles), maximization (white square), and numerousness (black triangles).

Though the analysis of the first 12 trials evidenced that no subject maximized its payoff in all conditions since the early beginning of the experiment, three of the four maximizing subjects significantly preferred token B when presented with 1B vs. 1A and 1B vs. 2A, and one and two subjects significantly chose tokens A when presented with 1B vs. 4A and 1B vs. 5A, respectively. The most problematic condition was 1B vs. 4A, in which only one capuchin

preferred tokens A over token B above chance level.

In a final experiment, the capuchins which preferred token B regardless of the comparison presented (black circles in Figure 4.10) were excluded and the remaining 6 had to maximize their payoff when two tokens B were paired with several tokens A. Capuchins faced three types of new comparisons (2B vs. 4A, 2B vs. 5A, and 2B vs. 6A), each presented twice; moreover, three types of familiar comparisons (1 B vs. 5A, 1B vs. 4A, and 1B vs. 3A), each presented twice, were interspersed in each session for a total of 12 trials presented in a pseudo-randomly order. The results showed that *only one subject* maximized its choice by selecting the token B when this was advantageous (i.e. when two tokens B were presented against four or five tokens A), and selecting tokens A when 4 or 5 tokens A were presented against one token B. A second capuchin behaved similarly, but her performance did not reach statistical significance for B choices in the 2B vs. 5A comparison. A third was at chance level in all conditions and the remaining three subjects developed a preference for token B regardless of condition.

In sum, this study may be said to have demonstrated that capuchins' performance in relative numerousness judgment tasks (RNJs) showed a similar pattern when they choose between different quantities of food and between different quantities of tokens that are associated with the corresponding amount of food. That the ratio between arrays influenced capuchins' performance strongly supports the idea that they process the two problems by relying on the same system for numerical representation, namely the analogue magnitude system or core system 1 (e.g. Feigenson et al. 2004, Nieder 2005). While there was large individual variation, two of 10 subjects seemed to be able to discriminate between two types of tokens on the basis of their different hedonic value and could estimate token quantities to maximize payoff. From an evolutionary point of view, cognitive mechanisms allowing animals to compare quantities may be adaptive, particularly in species that forage on patchily distributed sources whose individuals most of the time may be unable to compare food patches (Olthof et al. 1997, Call 2000). Action enables the development of relational mental activity (Fragaszy et al. 2004), and since capuchins have an extraordinary tendency to engage the world through direct physical contact with objects and to combine objects, they are likely candidates for the emergence of relational thought.

These are significant findings, but where we (the authors of this report) diverge from our colleagues who performed the study (cf. Addessi et al 2007) is on (a) whether these results show that capuchins are capable of "using symbols" and (b) whether the analogue numerical capacity demonstrated can be truly regarded as "representational" and involving "abstract thought" (D13: 33-34). The reason for this is that our colleagues adopted a different definition of "symbols" from the one that we find most appropriate.

> "In conclusion, our study evaluated the use and understanding of symbols in nonhuman primates by means of a token exchange paradigm. Tokens have a symbolic nature since they are objects that arbitrarily stand for something else without having any iconic relation to their referent (i.e., the different quantities of food items)." (D13: 33).

The fact that there is no iconic relationship between the token and the different amount of food rewards is not sufficient to guarantee their symbolic nature. In particular, the association that was learned was of an *indexical* nature: token and reward were for the capuchins temporally contingent. What evidence is there that this lead to the formation of a common experimenter-subject *convention* of the relationship? Visalberghi (p.c.) suggests that such "common knowledge" was inadvertently shown when, by mistake, the experimenter gave the wrong amount of food items in exchange for token B, e.g. two instead of three items. In those cases, the animals "protested", demanding the remaining reward. However, no experiment was carried out to test this explicitly, e.g. to see whether those subjects which protested more were also those who performed best in the maximization-of-profit experiments. Given the

large inter-subject variation in these experiments, the most parsimonious explanation remains that the capuchins performed on the basis of indexical (association-based) knowledge. This is indeed an ecologically useful capacity for a highly dexterous species such as capuchins. It is also a "flexible" one, likely to lead to different associations in different individuals, depending on prior history and "smartness". Methodologically, the most important conclusion concerning symbol use from the study is that symbols are quintessentially *social* phenomena, and experiments aiming to establish their existence need to use a paradigm demonstrating capacities in social cognition and communication, rather than object manipulation and quantity choice.

4.4.2 Tolerance and food sharing in great ape mother-infant dyads

Daniel Haun and Josep Call investigated whether there are species differences across all great ape genera in levels of tolerance and willingness to share food (D19). It is well known that interactions with their mothers strongly impact infants' socialization (Trevarthen, 1979). Higher levels of tolerance and cooperative behaviour between mothers and infants can be seen as both a reflection of a capacity for (primary) intersubjectivity, and as a scaffold for its further development. Ultimately, tolerance and cooperation can be seen as prerequisites for the establishment and successful cross-generational maintenance of cultural traditions. This study, therefore aimed at providing insights into species differences in early socialization across the great apes, which might be related to differences in the ontogeny of social-cognitive abilities detected in later life. The study involved six great ape mother-infant dyads: 2 chimpanzees, 1 bonobo, 1 gorilla and 2 orangutans, all from the Wolfgang Köhler Primate Institute in Leipzig.

In a first experiment Haun and Call made desirable food available either (a) inside the enclosure, so mother and infant could both access it or (b) outside, accessible only to the infant. In this latter situation, infants still had to bring the food inside to eat. Thus in both cases the infants' success greatly depended on their mothers' tolerance. The researchers coded the food intake of infants (the average number of grapes eaten by the infant in both conditions) and the positive (e.g. begging and food sharing) and negative (e.g. aggression) communicative interactions between mothers and infants. It was found that chimpanzee as well as the bonobo mothers showed extended tolerance towards their infants. While infants obtained little of the food available to both mother and infant, they obtained almost all the food only accessible to them. This implies that the mothers tolerated their infants eating even after the infant had brought the food inside the cage. There was hardly any aggressive interaction between them and in rare occasions the mothers would beg from their infants. In gorillas and orangutans, the picture was completely different. Mothers exhibited little tolerance towards their infants if those had food in their possession. Both gorillas and orangutans attempted, and most of the time succeeded, to take the food from their offspring, in case of orangutans even after the infants have taken the food into their mouths. Furthermore infants usually stopped reaching for food, in which case the mothers often attempted to use their infants as tools to obtain out-of-reach food items.

In a second experiment Haun and Call gave desirable food to either the mother only or the infant only. The possessor had the option to share or avoid sharing. Every dyad was run through four sessions on four different days. It was coded whether either mothers or infants would share food, and if so, if the possessor would actively support sharing or merely tolerate theft. No instances of active sharing of food items were found. However the researchers recorded instances of *tolerated theft* and found them to be more frequent in chimpanzee and bonobo mother-infant dyads than either gorilla or orangutans.

In conclusion, considering their cladistic relationships, the increased levels of tolerance and sharing in Pan (chimpanzees and bonobos) relative to the other non-human great ape genera suggests an evolutionary innovation in the great ape lineage leading to greater

tolerance and food sharing – which may reflect a higher degree of empathy, i.e. a form of intersubjectivity – after the split between the Pan and Homo lines on the one side, and the Gorilla and Pongo lines on the other. In other words, non-human great ape species that are most closely related to humans showed greater tolerance and were more likely to share food more than others.

4.4.3 Mutual gaze and physical contact in human and ape mother-infant dyads

Mutual gaze (MG) is a basic form of experience-sharing between infants and their mothers (and other care-givers). At least in Western cultures human infants participate in extensive bouts of eye-contact with their mothers more or less from the first days of life. Together with co-occurring vocalizations, beginning from about 2 months, such interactions have been regarded as a conspicuous and developmentally significant form of *primary intersubjectivity* (Trevarthen, 1979; Reddy, 2003). Such interactions play an important role for the emergence of understanding others as attentional and intentional beings (subjects), as well as one's own "sense of self" (Stern, 2000 [1985]).

In both (early) "dyadic" and (later) "triadic" interactions (the first type involving only infant and caregiver, while the second also include an external object), is has been claimed that human infants engage in considerably more and longer bouts of mutual gaze than apes (Tomasello, 1999). Such differences point to a basic *phenotypic* difference between human and ape intersubjectivity. However, whether this difference arises from genetic, or environmental-ontogenetic factors is not yet clear. It has been recently argued by Bard et al. (2005) that chimpanzees and humans engage in approximately the same rates of mutual gaze in early infancy (1-3 months), and that differences are due mainly to cultural/environmental rather than genetic factors. The authors compared mother-infant dyads from two groups of chimpanzees: at Yerkes National Primate Centre (n=8), where a mean of 12 instances per hour was measured, and at the Primate Research Institute in Kyoto (n=3), with a mean of 22 instances of mutual gaze per hour. The mean of approximately 17/hour was said to be comparable to that found in children, 18-20/h (Ling & Ling, 1974).

Interestingly, Bard et al (2005) found significant differences between the two groups of apes not only in terms of mutual gaze, but in terms of activities involving physical contact and in particular "cradling", which at 3 months of age was reversely correlated with mutual gaze: 71% of the observed time in Yerkes and 40% in Kyoto were spent by the mothers "cradling" their infants. Since a similar negative correlation between gaze-contact and physical-contact has been observed when comparing Western and traditional cultures (the first are more like the Kyoto apes, the second more like those at Yerkes) the authors concluded that any differences between human and apes patterns of gaze, previously attributed to genetic factors, are rather the result of group and culture-specific interactions.[13] Still, even these findings, including the unusually frequent rate of mutual gaze episodes in the Kyoto apes (who had experienced extensive interactions with human care-givers and researchers before giving birth), confirmed that ape dyads engage in much *shorter* episodes of mutual gaze than human dyads. Furthermore, since Bard et al. did not perform any coding of human mutual gaze using the same criteria as for the apes it is difficult to be completely convinced by the authors' claims of there being no essential species differences. Finally, even Bard et al. accept that there are differences in the "likely developmental course after 3 months" (: 621).

The goal of a study performed by Zlatev and colleagues (D19) was therefore to attempt to "replicate" the results from Bard et al (2005), comparing 5 ape dyads, and 5 human dyads, with the infants being somewhat older than those in the Bard-study, but still in the period of

[13] These claims are rather surprising, since standard text books in proxemics would claim that there is both less eye contact and less touching in Japanese culture as compared with Western culture.

primary intersubjectivity prior to 9 months. The main questions were (a) whether we would find similar rates of mutual gaze between the groups as a whole, (b) if not, could differences be accounted for with corresponding differences in physical contact (as between the two groups of apes in the Bard-et-al-study) and (c) whether there were large *intra-group* variation – due to species, environment and parenting style differences.

The ape data included three chimpanzees, one bonobo and one gorilla dyads. All ape infants were between 4.5 and 9 months at the onset of recording, apart from one of the chimpanzees (Lobo), who was 19 months. For the human data, 5 native Swedish families were recruited in Lund. For both groups, between 2 and 3.5 hours of spontaneous interaction data was collected, giving a total of about 14 hours per group. Removing "nonobservable" time (see below), there was 8.5 hours for the apes, and 12.5 hours for the children. The data was independently coded by two coders (and cross-checked by a third) for the following categories:

- **mutual gaze**: when (a) the eyes of both subjects were visibly directed at each other, or (b) when one of the subjects looked at the face of the other, and the latter made a movement of the head or body, resulting in a position facing the directed gaze of the other subject
- **physical contact**: when the mother – and in one occasion the father – held, hugged, cradled or nursed the infant (i.e. more than simply a touching between the two)
- **nonobservable**: when the head of either the infant or the mother (or both) where out of view

The results showed conspicuous differences between the species (Table 4.2), with a highly statistically significant difference between the groups. The low rates of MG in the ape data found in a preliminary report (D9) were corroborated after the elaborate procedure involving two, and in some cases three, more coders. The rates of MG for the bonobo were considerably higher than those for the other species (especially contrasted with the gorilla infant at the same age), but still much lower than the figures presented by Bard et al (2005). On the other hand, while the ape mothers held or otherwise engaged in a strong form of "physical contact" with their infants, on average more than the human mothers, 74% and 56.6% respectively, the differences were not significant. Furthermore, there were no negative correlations between physical contact and mutual gaze, for individual dyads or for the groups.

Table 4.2. Mutual gaze in instances per hour (visible time) and average duration for the ape and human dyads.

Apes	MG/hour	Length (sec)	Children (fictive names)	MG/hour	Length (sec)
Luiza (bonobo, 7m)	4.38	1.02	Kate (5;0m)	33.96	3.77
Kara (chimp, 5m)	1.45	0.61	Mårten (5;18m)	68.93	2.89
Kofi (chimp, 4.5m)	2.23	0.75	Walt (5;22m)	34.74	4.93
Lobo (chimp, 19m)	1	0.87	Anna (7;6m)	14.77	2.29
Sammi (gorilla, 9m)	0.67	0.62	Eric (8;11m)	22.05	2.11
Mean: 8.9m	1.94	0.82	Mean : 6;11m	34.89	3.33

While displaying huge intra-group variation (from 15/h to the rather incredible 69/h), MG episodes in the human dyads were as a whole nearly 18 times more frequent and 4 times longer than what was found for the apes. It is difficult not to see these differences as reflecting qualitative distinctions. However, it is not clear what the difference between these results and those of Bard et al. are due to. Age is a possibility, but the youngest ape infants

(Kofi and Kara) were not those with the highest rates of MG, but rather the 7 month old bonobo Luiza. Given the relatively low number of subjects in the study and the presence of only one group for apes and humans each, we cannot draw any strong conclusion concerning species-specificity. The study does however raise doubts concerning the rather strong conclusions of Bard et al. More research is needed to establish the nature of the observed phenotypic differences between ape and human mutual gaze, but even on a descriptive level, the study supports the hypothesis that human mother-dyads (in Western middle-class contexts) are particularly disposed towards primary intersubjectivity.

4.4.4. Perceptual intersubjectivity in human and ape mother-infant dyads

The three levels of perceptual intersubjectivity (cf. Section 2) were operationally defined and attested in a study comparing human and ape mother-infant dyads (D7; Zlatev et al. 2008). The definitions of the levels, limited to visual attention, shown in Table 4.3 were used for the study.

Table 4.3. Operational definitions of the levels of perceptual intersubjectivity used for the empirical study

Level	Term	Operational definition
1	Synchronous PI	INFANT: GAZE TO NEW T (+ NON-COMM. REFERENTIAL BEHAVIOUR TO T) ADULT: GAZE TO T
2.1	Coordinated PI - simple	INFANT: GAZE TO NEW T + COMM. REFERENTIAL BEHAVIOUR TO T ADULT: GAZE TO T
2.2	Coordinated PI - complex	INFANT: GAZE TO NEW T + COMM. REFERENTIAL BEHAVIOUR TO T + GAZE-TURNING TO ADULT ADULT: GAZE TO T
3	Reciprocal PI	INFANT: GAZE TO NEW T + COMM. REFERENTIAL BEHAVIOUR TO T + GAZE-TURNING TO ADULT + MUTUAL GAZE WITH ADULT (WITHIN IN OWN TURN) ADULT: MUTUAL GAZE WITH INFANT + GAZE TO T

The main hypothesis was that these levels correspond to *developmental* and possibly also *evolutionary* stages, i.e., they will be observed to different degrees in different periods of children's development, and will differ between apes (and possibly the common ape-human ancestor) and human beings. This hypothesis was tested in an empirical study using observational data from 3 apes (two chimpanzees and one bonobo), aged 7-19 months and 7 children (from Sweden and Thailand) aged 12-18 months.

Analysis was performed by two independent coders, and in case of uncertainty, a third coder was consulted until consensus was reached. Unclear examples were excluded. This resulted in a total of 190 PI episodes. The results strongly suggested that (asymmetric) PI is much more frequent in human than ape infant-adult interactions and consequently typical for human beings. In approximately 2 hours and 18 minutes of data, the apes engaged in only 5 asymmetric PI episodes. In contrast, the 2 hours and 8 minutes of human data contained a total of 185 instances. RPI (Level 3) episodes were not limited to the interactions of the 18-month old children, but they were proportionally more frequent than for the two 12-month olds. CPI episodes (Level 2) occurred in the data of both groups of children, but where altogether absent (along with RPI episodes) in the ape data. All 5 instances of asymmetric PI episodes initiated by the infant apes were cases of SPI (Level 1), cf. Figure 4.11.

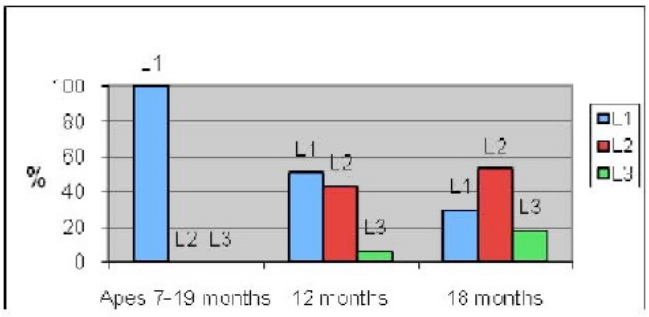


Figure 4.11. Percentages of asymmetric PI episodes by level (1, 2 and 3) in the three sets of data. Total number of episodes: 5 for apes, 50 for 12-month old children, and 135 for 18-month old children

The results of the study supported the hypothesis that the different levels of PI correspond to developmental and evolutionary stages. The three ape infants studied (in the four data points) and their interacting adults engaged in a surprisingly low number of PI episodes, and all of these were of the simplest type, Level 1 (SPI), which did not involve intentional communication. Even for the adult, this type does not necessitate the presence of more complex processes than *attentional contagion*. The complete absence of any higher level of PI episodes in the ape data, irrespective of the differences in the ages of the ape infants (7-19 months) supports the analysis of SPI as being qualitatively different from the higher-level types, and indicates a corresponding difference between Pan and Homo.

At the same time, we need not interpret this as a matter of inability of apes to engage in more complex types of intersubjectivity, since it is known from previous research that in captivity (adult) apes do engage in intentional communication with human subjects (Leavens & Hopkins, 1999). Nevertheless, the differences were so conspicuous, that they seem to reflect a *qualitative difference in the nature of ape and human social interactions*: human infants (and young children) engage in communicative referential behaviour on a regular basis, while ape infants do not. Furthermore, since the differences between the pre-verbal and "just-verbal" children were relatively minor (as reflected in the minor differences in CPI (Level 2) in Figure 4.11 for the two groups) this seems to be a feature of human social interactions that is independent of and more basic than language.

4.4.5 Differences in norms concerning pointing in Sweden and Thailand

Adult gestures, including (predominantly) deictic gestures such as pointing are influenced by specific societal norms. Kita and Essegbey (2000) showed how a "taboo" on the use of the left hand in Ghana for many forms of social interaction influences gestural practice in particular ways. Thai culture has an explicitly formulated taboo against pointing with or manipulating objects with one's feet. But there are also other less explicit norms involving pointing, which are more difficult to verbalize, and rather a matter of the culture's "habitus" (Mauss 1934; Bourdieu 1977). With respect to bodily communication, we can find recommendations such as the following being addressed to visitors to Thailand: *"Loud voices, calling attention to yourself, pointing at people or things, throwing things, and making big hand gestures, all seem graceless to the Thai sensibility." "Also be moderate with pointing your finger and avoid physical contact."*

An important question is when and how such *normative conventions* concerning the use of gestures and other similar forms of bodily communication – both the more explicit, and the more implicit ones – are acquired by children. In order to investigate this, an elicitation-based study was conducted, involving 16 children in Lund and Bangkok, divided in two groups of pre-schoolers: 7 Swedish (average age 4;8) and 8 Thai (average age 4;7) and 2nd grade children: 9 Swedish (average age 7;10) and 8 Thai (average age 8;8). (D22). Each child was

asked to sit beside a low table on which three fluffy toys (a dog, a rabbit and a teddy bear) and three plastic plates (with a spoon, a fork and a knife) were placed. On the lateral side of the table sat the interviewer (Int), and near one of the corners of the room stood the main experimenter (Exp) near the camera. A question-answer session, along with a "hiding game" was performed, using questions and contexts carefully chosen in order to contrast a number of different *conditions*, the most important of which were:

(1) pointing towards a visible inanimate object vs. the place where it was "hidden"
(2) pointing towards a visible (adult) person vs. the direction in which he left
(3) pointing towards a visible inanimate object vs. visible (adult) person

It was hypothesized that if the Thai children found pointing in general "graceless" they should do so for all conditions, and thus point less often than the Swedish children in general (H1). However, if the children found it specifically more inappropriate to point to present people than to present objects, this would be reflected in condition 3 (H2).

The results showed a clear developmental tendency with respect to the norm "do not point at people" for the Thai children: while the younger group pointed nearly as much to Exp as to the toy, there was a remarkable difference, both quantitatively and qualitatively for the 8-9 year old children (Figure 4.12). No such tendency was present for the Swedish children (Figure 4.13).

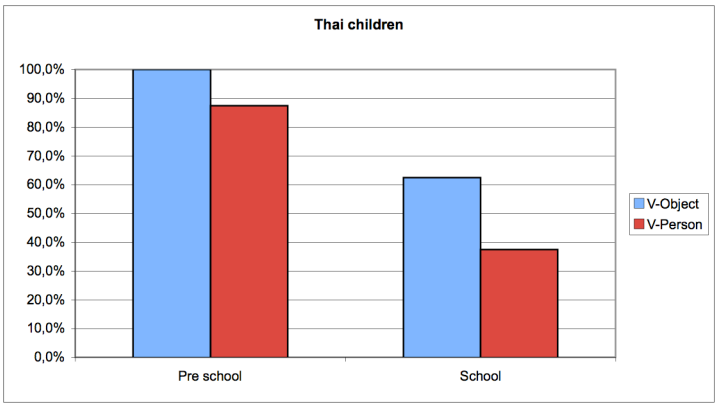


Figure 4.12. Index finger pointing with respect to the contrast person/object for the two age groups of Thai children

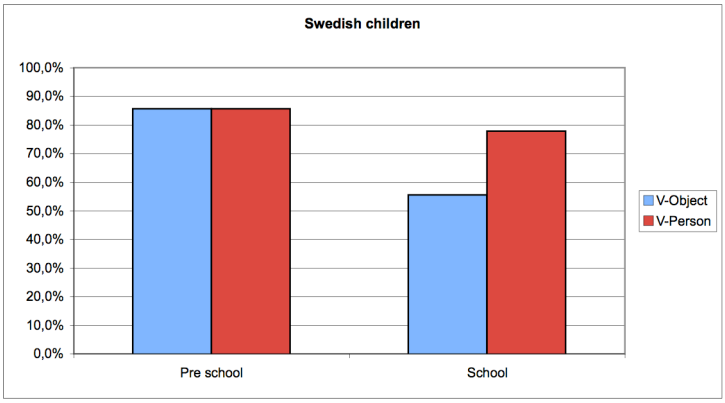


Figure 4.13. Index finger pointing with respect to the contrast person/object for the two age groups of Swedish children

The results supported clearly H2: only for the Thai children, and especially for the older group: there were 5 index finger points to the present dog vs. 3 to the present person. Furthermore, *not one of these three was a prototypical outstretched hand point*, but involved a gesture in which the elbow was held too close to the torso, with the finger pointing briefly in the direction of Exp. In the remaining 5 (of 8) cases the children either answered verbally ("the person close to the camera") or using a "polite" open-hand offering-like gesture. On the

other hand, of the two older Swedish children who did not use an index-finger pointing gesture in reply to (d) "Who moved them?" one used a rather unusual thumb-pointing gesture and only one used a "polite" offering-like gesture with a flat hand palm up. H1 was not supported by the results from the pre-school groups, but finds support for the older children (Figure 4.9). Figure 4.13 would appear to show that this is so also for the Swedish children, but this is somewhat misleading: when they did *not* use index-finger pointing, they grabbed and sometimes even picked up the toy, while the older Thai children either described the location verbally, or used a careful open-hand indicating gesture.

In sum, there were few differences between the Swedish and Thai pre-school children, while those between the school children were obvious. This indicates that acquiring the norm according to which index finger pointing is "impolite" is a relatively late achievement, and one that possibly rests on the teaching of explicit rules. However, as with the case of the even stricter taboo against left-hand pointing in Ghana, the application of such a norm is not "absolute" but allows for substantial individual and contextual variation. Replacement strategies were observed in the older Thai children (whole hand, "offering" gestures) not unlike those in the adults in Ghana who could point with both hands to avoid the taboo.

Whether there is a Thai norm against index finger pointing *in general* is less clear, both from the analysis of the corpus (Section 4.5.1) and from the elicitation study. (All of the Thai caregivers used index finger pointing in addressing their children, and no noticeable attempts to teach such a norm to the children were observed.) At the same time, it is clear that in a fairly formal school-context, pointing was perceived as inappropriate by the majority of the 8 year-old Thai children, mostly towards the visiting adult, but in many cases also towards the toy.

4.4.6. Linguistic conventions and the experience of motion

Jordan Zlatev and Johan Blomberg, from Lund, in collaboration with Caroline David from University of Montpellier 3, carried out a study with a twofold goal (cf. Zlatev, Blomberg and David in press). The first was to provide an experientially-based classification of *motion situations.* The second goal was to use this taxonomy of motion situations in addressing the questions of how different languages *conventionally* express motion, and if such differences imply differences in conceptualization.

Talmy (1985, 2000) considered the 'presence of motion', or *motion* with a small letter, along with the conceptual components *figure, ground, path* and *manner/cause* to be building blocks of a 'motion event', and depending on the way they are mapped to different constituents in the clause, formulates the basis for his well-known motion-event typology, shown schematically in Figure 4.14, with example sentences from English (a satellite-framed, or S-language) and French (a verb-framed, or V-language). This typology has been claimed to be exhaustive, i.e. that every one of the world's languages can be categorized as being, predominantly, an S- or a V-language.

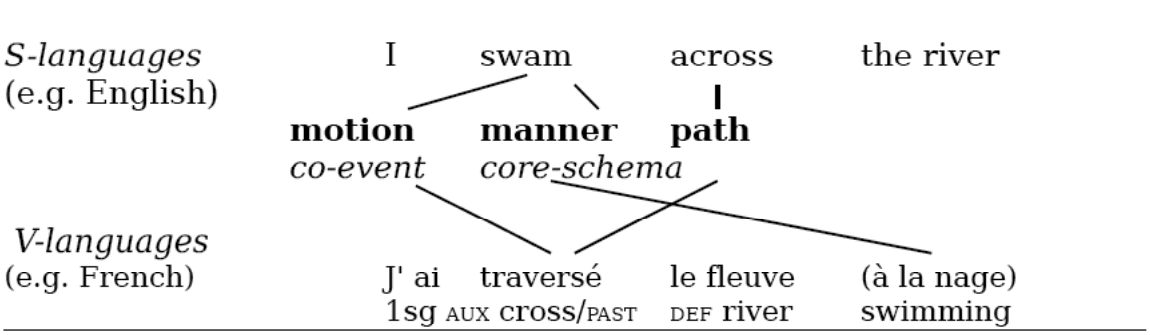


Figure 4.14. Different mapping patterns between the conceptual components of motion events and parts-of-speech in satellite-framed (S) languages and verb-framed (V) languages

However, it has become increasingly clear that this binary typology cannot do justice to the complexity found in the world's languages: either more 'exotic' ones such as Tzeltal (cf. Brown 2004), or more familiar ones such as Russian (cf. Smith 2003), as many of the contributions to the volume edited by Strömqvist and Verhoeven (2004), e.g. Slobin (2004) testify. In some of the group's previous work (Zlatev and David 2003; Zlatev and Yangklang 2004), it was documented how Thai, and by extension other similar serial verb languages, constitute a distinct 'third' type. For example, Thai resembles V-languages in some respects (e.g. path expression by a main verb), S-languages in other respects (e.g. manner expression by a main verb), while in yet other respects it resembles neither (e.g. by having a separate 'slot' in the serial verb construction for path+manner conflating verbs), cf. Zlatev and David (2003) for discussion.

The basic, and yet unresolved, question however, remains 'What is motion?' and correspondingly: 'What is a motion event?' From the perspective of the analysis of (the invariants of) experience – *phenomenology* (cf. Husserl 1909), motion as such can be defined as the experience of *continuous change in the relative position of an object (the figure) against a background*, in contrast to stasis – where there is no such change – and in contrast to a *dis-continuous* change, as when a light suddenly lights up in position A, 'disappears' and then appears in position B. As well-known, however, if the time fragment between the two discrete events is small enough then an observer will actually see the light as *moving* from A to B, in a continuous manner. Thus, motion is 'in the eyes of the beholder'. Note that 'continuous' is here meant to exclude from the definition of motion such events as disappearing at one place, and reappearing at another, as in a Star Trek case of 'teleportation', which may be in the sphere of the imaginable, but not in the ordinary human lifeworld. It does *not* exclude instances of rather abrupt types of motion, e.g. jumping, blinking, breaking or other similar 'punctual' events.

Furthermore, note that motion 'from A to B', i.e. *relocation* (Smith 2003) is not a necessary characteristic of a motion situation. First, the light could waver around A, and then there would be no change in its average position and thus there would be 'self-contained' motion in Talmy's terms. Second, the figure could be moving along a vector in an open-ended way, for all eternity perhaps – and hence there need not be any B to relocate to. Third, the figure's motion can be either spontaneous or caused by an external source. Thus, we have *three different parameters* according to which motion situations can vary, quite independent of their representation in language. The independence of the three parameters yields the 8 types of motions situations illustrated in Table 4.4, with schematic representations in English.

Table 4.4. Illustration of the expression of 8 motion situation types in English; F = Figure, LM = Landmark, A = Agent, View-C = Viewpoint centred, Geo-C = Geocentric, Obj-C = Object centred Frame of Reference

	-CAUSED	+CAUSED
+TRANSLOCATIVE +BOUNDED	F goes to LM	A throws F into LM
+TRANSLOCATIVE -BOUNDED	F goes away (View-C) F goes up (Geo-C) F rolls forward (Obj-C)	A takes F away (View-C) A pushes F upward (Geo-C) A pushes F forward (Obj-C)
-TRANSLOCATIVE +BOUNDED	F breaks (up/down)	A breaks F (up/down)
-TRANSLOCATIVE -BOUNDED	F waves	A waves F

While Talmy made the notion of 'motion events', or rather – translocative situations – into a popular subject for typology, it was Slobin (1996) who brought the subject to the attention of neo-Whorfian research on linguistic relativity. According to one of Slobin's formulations, it may even be a mistake to look for language-independent taxonomies of situations, since:

> The world does not present 'events' and 'situations' to be encoded in language. Rather, experiences are filtered through language into verbalized events. A 'verbalized event' is constructed online, in the process of speaking. (Slobin 1996: 75)

But at the same time, Slobin's famous 'dynamic' formulation of the Whorfian program, known as *thinking for speaking*, only concerns the 'special kind of thinking […] that is carried out, on-line, in the process of speaking' (Slobin 1996: 75) and is therefore different from Whorf's (1956) notion of 'habitual thought', according to which language should have much more pervasive effects (cf. Blomberg 2007). Methodologically, Slobin (1996, 1997, 2003) concentrated on differences in the 'rhetorical style' of speakers of V-languages such as Spanish and S-languages such as English – as something that could be explained by the languages' different ways of expressing, above all, the concepts Path and Manner. For example, due to the optional expression of Manner in V-languages (see Figure X.1), their speakers were found to express Manner less often and preferred to give more static descriptions in which the Figure's motion could be inferred from the 'scene setting' and the result of the motion, while S-languages induced descriptions in which the events were presented more dynamically, with more elaborated representations of the Path. But, as pointed out by Pourcel (2005), Slobin's research gives little support for strong relativistic effects in *the categorization of experience as such*, i.e. even when 'thinking for speaking' is (apparently) not involved.

A number of other studies have attempted to demonstrate such effects using, among other methods, a classic task for studying categorization in an (apparently) non-linguistic context: *forced-choice similarity judgments*. The general method, used with various modifications, in all of these studies is to use *triads* of representations of motion situations: a target situation is presented along with two alternatives, where one differs from the target with respect to Path and the other with respect to Manner, and the subject is asked which of the two 'is most similar' to the target. The general reasoning is that if language impinges on categorization, then speakers of a V-language should be predisposed to prefer 'same-Path' rather than 'same-Manner' to a greater extent than speakers of S-languages, where both components are expressed equally easy. An exception to this line of reasoning was offered by Papafragou, Masely and Gleitman (2002), who suggested an alternative basis for a linguistic effect that actually runs in the opposite direction: since Manner is often expressed in a non-obligatory constituent in a V-language, when it is expressed, it would be 'foregrounded' and thus achieve more *semantic salience* (Talmy 1985) than in an S-language where it is

expressed by an obligatory constituent, such as the main verb. Papafragou, Masely and Gleitman (2002) compared among other things the categorization of triads (using static pictures) by speakers of Greek (assumed to be a V-language) and English (an S-language) and despite differences in the linguistic descriptions that followed the predicted patterns (along the lines of Slobin's research), they found no bias for either Path or Manner-based judgments in either group, and thus argued against the presence of any Whorfian effect on motion event categorization.

In previous research, Zlatev et al. used the *Event Triads* elicitation tool, developed at the Max Planck Institute for Psycholinguistics, Nijmegen (Bohnemeyer, Eisenbeiss and Narasimhan 2001), which was created to investigate biases for Path or Manner in forced-choice similarity judgments, and compared similarity judgments of Swedish, French and Thai speakers. Contrary to our expectations, it was not the Swedish, but the Thai group that had the largest proportion of same-Manner choices: the difference between the Thai group on the one side, and the French and Swedish groups on the other was statistically significant, while that between the French and the Swedish groups was not. However, when the results when we divided the 12 test triads according to the three types: FROM/TO, OUT/INTO and VERTICAL, the classification of the Vertical unbounded translocative situations for the French group differed significantly from the other two types of situation, while there where no such differences for the other two languages. Furthermore, this difference was correlated with different linguistic descriptions for the Vertical, compared to the other two types of motion situations in the French group.

In a new study, carried out within SEDSU, Zlatev et al. modified the Event Triads elicitation tool so that two groups of 12 Swedish and two groups of 12 French subjects participated: Group 1 for both languages performed the similarity judgment as in the original Event Triads tool, whereas for Group 2 there was a break after the first segment and the participant was asked to 'describe the film just seen', after which the second segment was shown and the participant was asked to make the similarity judgment. The number of distracters was decreased rather drastically from 38 to 8, leaving the total number of triads per participant to 20, where each first segment was described by all participants: for Group 1 after the similarity judgment task was competed, and for Group 2 prior to each judgment. In this way we could investigate possible correlations between the descriptions and the choices not only on a type-by-type basis (as in Studies 1 and 2), but also on a triad-by-triad (instance) basis.

The results were quite interesting: Whereas the similarity judgments for Group 1 (post-choice description) were similar to those in Study 1 and practically identical for the two languages, i.e. a preference for same-Manner choices, the situation was completely reversed for Group 2 (pre-choice description), with a surprisingly strong bias for same-Path choices, as shown in Figure 4.12 for both the French and Swedish groups. The difference between Group 1 and Group 2 was extremely significant. Furthermore, there was a stronger Path/Direction bias for SG2 than FG2, which was also significant.

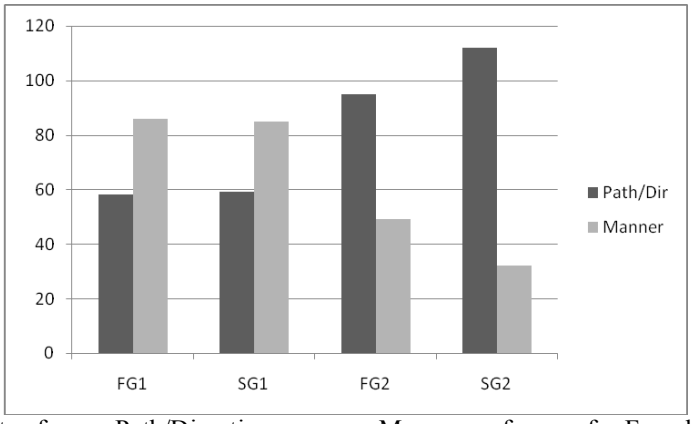


Figure 4.15. Total results of same-Path/Direction vs. same-Manner preference for French (FG1) and Swedish (SG1) Group 1 (post-choice description) and French (FG2) and Swedish (SG2) Group 2 (pre-choice description). Total number of choices is 144 per group

The researchers coded the linguistic descriptions for the presence of Manner expressions: Manner verbs such as *hoppa* and *sautille* ('jumps') and adverbials such as *snurrande* or *en roulant* ('rolling'), Path expressions such as *från* or *de* ('from') and *till* or *a* ('to') and Direction expressions such as *upp* ('up') or *monte* ('climbs') and *ner* ('down') or *descend* ('descends') and looked for correlations between the presence of these elements and the choices of the subjects. Surprisingly, there were few *positive* correlations: for SG2-Vertical and for FG1-Vertical. In qualitative terms, this means that if a subject had used a Direction expression, he was more likely to make a same-Direction than same-Manner choice. Thus, the results did not lend themselves to an explanation in terms of Slobin's (1996) thinking-for-speaking hypothesis. According to the latter, and the classification of Swedish as an S-framed and thus Manner-salient language and of French as a V-framed language, one would have expected a Manner correlation for SG2 a Direction/Path correlation for FG2. In fact, the only clear correlation was for SG2, and it involved Direction rather than Manner.

Applying the notion of *linguistic mediation* (Vygotsky 1978) to the triad studies allowed us to make sense of most of the results reported in the literature. First, due to the nature of the task, the similarity judgment task can be performed either more *directly* (i.e. using perceptual categorization) or more *mediatedly* (i.e. using external or internal speech). This can explain the results of both Gennari et al. (2002) and Finkbeiner et al. (2002), in which a typologically congruent bias was observed in the tasks where language was used either overtly or (apparently) covertly, but not otherwise. On the other hand, if Manner is a category which is (in general) more perceptually and conceptually *simpler* than Path, then tasks which induce categorization through less mediated processes, should bias for Manner rather than Path, and vice versa. We can thus explain the results of the study for both the Swedish and the French groups through a possible 'Vygotskyan effect' of language on the categorization of (translocative) experience: linguistic mediation yields an explicit 'parsing' of the components of a motion situation, and thus attention to more *abstract* components such as Path than to more perceptually immediate components such as Manner (or Direction). Such an effect appeared to be independent of typological differences between languages. At the same time, this interpretation predicts that if Manner is expressed in French, it will be more prominent *semantically*. Whether this would lead to a cognitive effect, however, is less clear: some previous research supported this, while the latest study did not.

In sum, the empirical findings suggest that the categorization of motion situations can be either more *direct* – and thus relatively unaffected by language – or more *mediated* (Vygotsky 1978), and that linguistic conventions can play a considerable role in the second case, but not necessarily in the first.

4.4.7. The conception of Time in Amondawa (and the West)

Vera da Silva Sinha, Chris Sinha, Jörg Zinken from University of Portsmouth and Wany Sampaio, from the Federal University of Rondônia (da Silva Sinha et al. in press) investigated the conceptualization and lexicalization of time intervals in an indigenous Tupi Kawahib culture of Western Amazonia: the Amondawa. Few studies have been published on the cultural and linguistic organization of temporal intervals in small-scale human groups whose traditional way of life is dominated by hunting, fishing, gathering and small-scale cultivation.

Using methods of linguistic and cultural analysis, as well as ingenious experiments, the researchers showed convincingly that the Amondawa time interval system is based not on countable units, but on social activity, kinship and ecological regularity. It does not permit conventional "time-reckoning" since the number system has only two numerals with a maximum combinatorial value of four. The Amondawa do not entertain cardinal chronologies such as ages of individuals, or ordinal chronologies such as yearly or monthly calendars. The researchers therefore claimed that the Amondawa, similar to another traditional culture, the Nuer (Levine 1997), live with a conception of "event time" rather than "clock time": activities are not fitted into a schedule governed by the clock or calendar, rather the temporal structure of life emerges from participation in daily activities.

This was contrasted to a view, predominant in Western thinking (both theoretical and of the more everyday variety) that Time is an autonomous, abstract conceptual domain, construed as a "plane" of thought-about, reflective experience, schematized in linear or cyclic terms, that is in some sense independent of the events that occur "in time". The authors referred to this conception as "Time as Such", with allusion to Kant. In contrast to universalist claims, including those which propose that the linguistic expression of time is universally structured via metaphoric mapping from the lexicon and grammar of space and motion, the study seemed to show that the domain of "Time as Such" is not a cognitive universal, but a historical construction based in social practice, semiotically mediated by symbolic and cultural-cognitive artefacts and subsequently entrenched in lexico-grammar. Importantly, the authors hypothesized that it is not language per se, but more general cultural practices (conventions) and artefacts, that play the decisive role in the construction of abstract concepts such as Time, and that this is what potentiates the widespread recruitment in the languages of the world of spatial linguistic resources for structuring the temporal domain: a hypothesis that would need to be tested against data from language change.

This important study was here summarized only briefly, because of time (sic!) limitations, but this does not adequately reflect its significance for the project. In particular, it was the only study that concerns explicitly what Donald (1991, cf. Section 5.1) calls the "third transition" into *theoretical culture* that is made possible by the invention of "ex-bodied" (external to the body) semiotic resources such as pictures and writing.

4.4.8 Conclusions

The studies described in this section were quite heterogeneous, but they were linked in that they either compared different species with respect to skills related to intersubjectivity, which can be viewed as pre-cursors to conventions (Section 4.4.2-4.4.4), or rather studied the effects of differing cultural and linguistic conventions on referential behaviour (4.4.5), motion categorization (4.4.6) and the conception of time (4.4.7). The study investigating the ability of capuchins to differentially associate tokens and rewards also investigated a type of precursor of a more general kind: the ability to establish and remember indexical relations.

Linking to the discussion in Section 2.2.3, we emphasized that conventions imply the ability to entertain common knowledge, and thus can be properly studied only in social contexts. The ability to form associative relationships may indeed be a precursor for this, but conventions in general, and conventional signs (symbols) in particular cannot be reduced to these. Hence the study which showed that (some) capuchins can associate tokens and food

items stably enough to use the tokens as a kind of mediator in their estimation of relative quantities, cannot be said to have established the presence of convention/symbol capacity in these distant relatives or ours. However, when combined with a more developed capacity to understand con-specifics as experiential subjects, this could have served as an important bootstrap for the learning of conventions.

The three comparative psychological studies of intersubjectivity investigate the subject within the domain where it is likely to have evolved in the primates: mother-infant interactions. Differences in tolerance and food sharing between, on the one hand, gorillas and orangutans and, on the other, chimpanzees and bonobos, support (however indirectly) the hypothesis that the human-chimpanzee ancestor already differed from the other apes in this respect. The studies showing differences in mutual gaze and perceptual intersubjectivity between apes (including chimpanzees/bonobos) and human mother-infant dyads further support the idea that skills for experiential sharing were further developed in our species. The fact that chimpanzees can in some contexts behave similarly to humans in these respects (though typically they do not) can be seem as evidence that intersubjective, proto-conventional skills such as mutual gaze, pointing (and other gestures) were within the common ancestor's "zone of proximal evolution" (Donald 2001), which on analogy with the Zone of Proximal Development (Vygotsky 1978) involves skills which can only be learned only "with the help of others".

Moving to the cross-cultural and cross-linguistic studies, we could see that different societal norms are indeed "internalized" by children, but at least in some cases (referring to other people), these require considerable amounts of time and enculturation. At the same time, as we will show in Section 4.5.2, some implicit understanding of these may appear much earlier. Since the question of the *emergence* of conventions requires a developmental methodology, the studies summarized in this section have not been able to address it properly. What the final three studies did show was that, predictably, knowledge and use of conventions has effects on pointing behaviour, the categorization of extra-linguistic situations and the formation of abstract concepts such as Time. However, while cross-cultural comparison was used in all these studies, only the first demonstrated differences related to different conventions in the two cultures (Sweden vs. Thailand). The study of motion categorization in Swedish and French speakers showed expected differences in the verbalization of experience, based on typological differences between Swedish and French. These differences, however, had little if any effect on the categorization of motion situations, while what was essential was whether (conventional) linguistic description was employed or not. This seemed to confirm the Vygoskyan position that human experience can be either relatively *direct* (unmediated by convention) or *mediated*, depending on the task and context, and a strong form of linguistic determinism with respect to perception and categorization appears unlikely (cf. Section 4.1). Also in line with this were the Amondawa findings that when it comes to more abstracts concepts, such as Time, varying cultural practices and semiotic artefacts have decisive roles.

5. Theories of semiotic evolution and development

As pointed out in Section 1, one the major goals of the SEDSU project has been to formulate a conceptually coherent and empirically founded novel theory of semiotic evolution and development. In this context, this means a theory that is consistent with the definitions of cognitive-semiotic concepts presented in Section 2, the Four Levels of Meaning framework suggested in Section 3, and the empirical findings summarized in Section 5.

This is a challenging task, so much more so because, at this "stage", with some of the empirical reports being only a few months old, it is still too early to be able to offer a more comprehensive theoretical synthesis. Also, given the number of different research groups and disciplines involved in the project, it is difficult to formulate a theory that would satisfy all our colleagues' theoretical commitments. Nevertheless, we find it important to think in terms having a wider scope, and there is no better way to do so than to try to formulate a theoretical synthesis, however provisional. Therefore in this section, we present two "theories of semiotic evolution and development" and relate these to what has been said so far. The first is the model of Merlin Donald (1991, 2001), which was inspirational from the onset of the project. Section 5.1 summarizes it in semiotic terms, pointing out how certain unclear aspects can be interpreted in light of our concepts and findings. Indeed, the main reason for taking up this theory here at the end is to show that, given a more thorough analysis of the terms of this model (as suggested by Sonesson 2006, 2007, in press), some of our "levels of meaning" rather naturally suggest themselves. Section 5.2 deals with the most overarching theoretical effort within the SEDSU project, the Mimesis Hierarchy model, developed by Jordan Zlatev, in association with colleagues at Lund University.

5.1 Semiotical reconstruction of Donald's model of evolutionary stages

Donald (1991, 2001) has posited several discontinuities in hominid evolution, all involving the acquirement of a distinct kind of memory, considered as a strategy for representing knowledge. The Tartu school of semiotics characterizes signs as memory devices and defines culture as collective memory. According to Lotman et al. (1975), material objects and information are similar to each other, and at the same time differ from other phenomena, in two ways: they can be accumulated, whereas for example, sleep and breathing cannot be accumulated, and they are not absorbed completely into the organism, unlike food, but remain separate objects after the reception. Projecting the double aspects of accumulation and communication suggested by the Tartu school onto Donalds' stage model, amounts to a first, essential, step in reconceiving the model in truly semiotical terms.

According to Donald's theory (though this is controversial), many mammals are already capable of *episodic memory*, which amounts to the representation of events in terms of their time and place of occurrence. The first transition (which Donald identifies with *Homo ergaster/erectus*) brings about *mimetic memory*, which antedates language and remains intact with its loss in aphasia. Mimetic representations are according to Donald "conscious, self-initiated, representational acts that are intentional but not linguistic" (ibid: 168). They are claimed to be required for such abilities as the construction of tools, miming, imitation, coordinated hunting, a more complex social structure and simple rituals. This stage thus in part seems to correspond to the attainment of the sign function (though Donald only notes this obliquely, in discussing the use of intentional systems of communication and the distinction of the referent from the mimetic expression). Yet, it should be noted that while all abilities subsumed in this stage seem to depend on iconic relations (perceptions of similarity), only some of them are signs because they do not involve any asymmetric relation between an expression and the content for which it stands.

Only the second transition, occurring with *Homo sapiens*, brings about language with its *semantic memory*, that is, a repertory of units that can be combined. This kind of memory

permits the creation of narratives, that is, mythologies, and thus a completely new way of representing reality. Importantly, however, Donald does not think semiotic development stops there, although further stages are no longer based on any biological changes. However, the third transition obviously would not have been possible without the attainment of the two previous ones. What Donald calls "theoretical culture" presupposes the existence of *external memory*, that is, devices permitting the conservation and communication of knowledge independently of face-to-face interaction between human beings. The onset of theoretical culture coincides, according to the theory, with the invention of drawing. For the first time, knowledge may be stored externally to the organism. The bias having been shifted to the visual modality, language is next transferred to writing. It is this possibility of conserving information externally to the organism that later gives rise to science (cf. Figure 5.1).

Considered as a form of memory, language/speech is clearly not as intimately bound to the body as mimesis: it is a conventional system which is shared among several individuals, and yet it is not open to third-person observation in the same way as organism-independent artefacts such a pictures, writing, and theory. Writing may be viewed as the transposition of speech to independent artefacts. The picture must be divorced from the bodies (and minds) of those making use of it. Students of prehistoric pictures such as White (2000) often suggest that creators of such works must have been capable of language. In fact, it is not easy to establish any clear-cut relation between linguistic capacity and the sophistication of the depictions. Not much can be concluded on the basis of the depictions having come down to us: even though pictures, by their nature, must have been made on material which conserves the markings on the surface, they might at first have been created on surfaces (such as sand or human skin) which only preserve them for a short time. In this sense, drawing may have debuted as mimetic memory, that is, as a skill for reproducing patterns, which, like gesture, is conserved, so to speak, in hands and arms. Even though evidence from developmental psychology seems to suggest that pictures are understood later than language (cf. Section 4.3.1), it is not clear whether this has something to do with the sign function, nor with them being external representations.

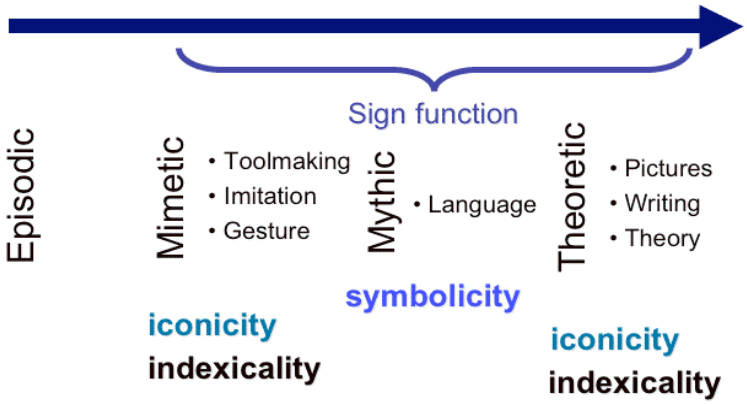


Figure 5.1. Donald's model of evolution related to the notion of sign function.

What is also made clear by regarding Donald's model from the point of view of semiotics, is that iconicity and indexicality precede the sign function, and then re-emerge again after the introduction of symbolicity with the sign function. This is not necessarily true, however, if the first picture creations were really comparable to gesture and were part of

mimetic memory incorporated into the body. Thus the emergence of icons and indices from iconicity and indexicality would occur *within* the mimetic stage, when going form one kind of imitation which is not as yet a sign to another kind which already has sign character: the first one would allow the propagation of skills such as tool-making from individual to another, and it would engender tokens which are of the same type, while the second one would be a gesture depicting in the air some kind of event in the world, i.e. pantomime.

Donald's evolutionary model is interesting also for what is fails to spell out. Imitation, or, more exactly, "representative imitation", is claimed by Piaget (1945) to be at the origin of the semiotic function (cf. Section 2.2.4). Donald places imitation within his second stage of human development, mimesis. To repeat, mimetic culture starts out with the emergence of "conscious, self-initiated, representational acts, which are intentional [i.e. voluntary] but not linguistic" (1991: 168). The examples of mimesis given by Donald are such things as gesture, dance, ritual, mime, play-acting, and (precise) imitation, but also tool use (or perhaps rather the social generalization of tool use) and skill. Curiously, Donald (1991: 170) claims to have derived his idea of mimesis from the literary theorist Eric Auerbach, who wrote a history of realist literature with this very title. It would have been more fitting to refer to the sense of the term "mimesis" in Antiquity, not perhaps as used by Plato to describe the relationship between perceptual reality and the world of ideas, but rather to one of the usages to which the term is put, mainly by Aristotle, as the representation of action by action, different from (verbal) narration or diegesis.

In his early book, Donald (1991: 168f) opposes mimesis to mimicry and imitation, both of which are said to be quite common in animals but lacking "a representational dimension". Though the import of this claim is not clear, it could be taken to mean that mimicry and imitation, in this sense, lack differentiation. In Donald's (2001: 260f) later book, however, "(precise) imitation" is an instance of mimesis. This would no doubt exclude the kind of neonatal imitation, discovered by Meltzoff, such as sticking out the tongue to one who does just that (Donald 2001: 264ff; Section 2.2.4). It is less clear whether Donald would follow Tomasello (1999) in making a distinction between the imitation of goals (called "emulation"), and the imitation of the means, which was initially claimed to be human-specific, although more recently its presence is recognized in at least some apes (Tomasello et al 2005). At first it may seem strange that imitating the goal is presented as being easier than imitating the means by which the goal is achieved. But no doubt it is less demanding to recognize the interest of the aim (getting the banana) than the interest of the steps requisite for realising the goal. At another level, it is like attending to the content, not the expression, of a sign. Indeed, it is an instance of quite ordinary Lifeworld behaviour.

The stage preceding the attainment of the language capacity, *mimesis,* requires memory to be located in the subject's own body. But, clearly, it can only function as memory to the extent that it is somehow separable from the body as such. The movement of the other must be seen as distinct from the body of the other in its specificity, so that it can be repeated by the self. This supposes another important semiotical distinction, that between *token* and *type,* preceding that of the sign function. That is, there must be some awareness (on whatever level) that the action which is repeated is, in *relevant* aspects, the "same" as that which serves as its model, while being different as an instance in time and space, as well as having numerous properties which are not relevant for its being a repetition of the first instance (this relationship picking up certain properties and ignoring others has been called the *principle of relevance,* both in linguistics, and in the phenomenology of Gurwitsch)[14]. Thus, for instance, when a student imitates the teacher's pronunciation of a word in a foreign language, the goal is not to render all aspects of the teacher's way of talking, but to realise the word as correctly

[14] This is similar to the use of this term by Sperber & Wilson (1996) in referring to a hierarchical organization of properties perceived in terms of more or less importance, but it is different from their use is supposing intersubjective rules for such hierarchies, rather than *ad hoc* situational cues.

as possible as given in the lexicon of the language. Imitation, in this sense, only indirectly refers to the act of the other as a way of attaining the common type.

Imitation also supposes a distinction between *Ego* and *Alter*. It will be noted that, in Piaget's conception, the differentiation of expression and content in the sign originally seems to coincide with the differentiation of self and other. This is only possible because, in the simple forms of mimesis, it is the motion of the self that serves as "expression" and the action of the other as "content". That it, imitation, in some form or other, seems to be that which permits the emergence of the sign function out of tokens and types.

In contrast, many kinds of mimetic acts mentioned by Donald, such as mime and play-acting and many instances of gesture, dance and ritual, are full-blown signs, in which an expression is clearly differentiated from a content for the user of the sign. The sign function may, of course, be of different kinds: the actor now on the scene represents Hamlet in Ellsinore during the Renaissance (play-acting); this group of children represent generic mothers, fathers, and children (symbolic play); a particular configuration of the hand represent victory as a general idea (gesture); etc.

One may wonder why tool use and skill are thought to be part of mimetic culture and not just "routine locomotor acts" or "procedural memory" which Donald (1991: 168) elsewhere takes pains to separate from mimesis. For Donald (1991: 171ff) this seems to be so because they comply with the criteria for mimetic acts in being "intentional" (that is, voluntary) "generative" (that is, analysable into components which may be recombined into new wholes), and "communicative" (or at least, as we shall see "public"), having reference ("in mimesis the referential act must be distinguished from its referent", that is, in our terms, there must be differentiation), standing for an unlimited number of object, and being auto-cued (produced without an external stimulus and therefore being the earliest form of "thinking"). Generativity is a property of many kinds of meaning (e.g. certain actions), which are not signs. However, it is not clear in what sense tool use and many other kinds of skill are "communicative", and therefore, in which way they have reference and stand for an unlimited number of objects. After introducing "communicativity" as a criterion of mimesis, however, Donald (1991: 172) states:

> Although mimesis may not have originated as a means of communication, and might have originated in a different means of reproductive memory, such as toolmaking, *mimetic acts by their nature are usually public* and inherently possess the potential to communicate. A mimetic act can be interpreted by others who possess a sufficient capacity for event perception. Given the pre-established primate capacity for event perception, the presence of mimetic skills would inevitably lead to some form of social communication (our emphasis)

This, however, is very different from imitation as a sign, which is what is realised by the actor, who presents his acts to a specific audience; it is even different from the child's symbolic play, which must be available to, and shared with, the other children. What we have here is, first, the extraction of a token from a type, which supposes treating the other as a "spectacle" (something to be observed); and second: the realisation of the tool act, which is not audience-directed, but can be made available to the public.

Thus, specific acts of tool use and some kinds of skilful performance are not in themselves mimetic (according to Donald's definition), because they are not communicative. However, they are "public", and they lend themselves to imitation – which leads to generalization of tool use and skill in society. This is where they become different from routine acts and procedural memory. They are socially shared (intersubjective). But this is only possible if the act can be separated from the unique tool user and transferred to another user. That is, the act as token must be abstracted to a type in order to be realised in another token. What is shared is the type, in other words the scheme of interpretation, which defines the principles of relevance. In this sense (not in the sense of reference), a single mimetic act may correspond to various events.

The conclusion from this discussion is that imitation (in one sense of the term) seems to

be a particularly important ingredient of the mimetic stage in that it naturally leads on to the acquirement of the sign function, first through non-conventional signs. In the so-called "mythic stage", conventional signs (symbols) first emerge. The use of external, organism-independent artefacts for representation may well also be an essential stage, following upon the earlier ones. It presupposes active perception, in that the principle of relevance singles out the features defining a type within a series of token events as well as the sign function. One particular form of it: the use of the picture sign, is firmly established in evolution only in historical time. However, this does not imply that understanding pictures as such is necessarily possible only after the attainment of conventional (linguistic) representations. As the SEDSU studies summarized in Section 4.3 showed, it was difficult to attest the presence of this capacity in non-human primates. However, at least one study, using the paradigm of do-as-I-do imitation, showed that a partially enculturated but non language-trained chimpanzee could reproduce actions given only photographs, including such representing a state prior to the completion of the action. Even if this may be interpreted as a form of stimulus enhancement, it seems to show that the chimpanzee could interpret the photographs as pictures, i.e. as representations that are similar to, but distinct from, the actions that they represent.

5.2. The Mimesis Hierarchy model of cognitive-semiotic evolution and development

5.2.1 Overview

As suggested in the previous sub-section, Donald's evolutionary model of semiosis has many attractive features, the foremost of which is perhaps its synthetic and comprehensive nature. At the same time, as also implied in the discussion in 5.1, there is much in Donald's model, and in particular in the concept of *mimesis* that has remained unclear. For example: What distinguishes mimesis from mimicry on the one hand, from signed language on the other? How exactly should we account for the semiotic differences between e.g. imitation and gesture, and did they evolve in any particular order? Are all aspects of mimesis under conscious control, or are some "sub-personal"? Which senses of the philosophically loaded (and ambiguous) notions of "intentionality" and "reference" are intended in the definition?

Empirically, the model is not really up to date with the evidence concerning ape capacities. It has, for example, been argued that some apes are in fact capable of some forms of imitation, e.g. (Custance, Whiten, & Bard, 1995; Whiten, 2000; Bjorklund & Bering, 2003), pointing (Leavens & Hopkins, 1999) and possibly even iconic gesture (Tanner, 2004). In terms of neuroscience, it is not evident how Donald's notion of a "mimetic controller" operating on mimetic memory relates to recent influential notions such as "mirror neurons" (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Arbib, 2005), "brain simulation" (Berthoz, 2000) and "shared representations" (Decety & Sommerville, 2003). Finally, given the somewhat vague nature of the concept, it has been applied quite differently to the analysis of human ontogeny (Nelson, 1996; Zlatev, 2003), making it difficult to evaluate suggestions that the model can account for a degree of parallelism between phylogeny and ontogeny.

In attempting to resolve these difficulties Zlatev and colleagues have proposed a series of related definitions of the concept of *bodily mimesis* (Zlatev, Persson, & Gärdenfors, 2005; Zlatev, 2005, 2007, 2008, in press), the latest version of which is the following:

> A particular bodily act of cognition or communication is an act of **bodily mimesis** if and only if:
> a) It involves a cross-modal mapping between *exteroception* (i.e. perception of the environment, normally dominated by vision) and *proprioception* (perception of one's own body, normally through kinaesthetic sense);

b) It is under conscious control and corresponds – either iconically or indexically – to some action, object or event, while at the same time being *differentiated* from it by the subject;
c) The subject *intends* the act *to stand for* some action, object or event for an addressee (and for the addressee to recognize this intention);[15]
d) Without the act being *conventional-normative*, and
e) Without the act dividing (semi)compositionally into meaningful sub-acts that *systematically* relate to each other and other similar acts.

This definition clarifies a number of issues on the relationship between bodily mimesis and similar but distinct phenomena. At the same time, it allows the formulation of a model of human cognitive evolution (and development) called the *Mimesis Hierarchy*. The model defines each successive stage through the clear attainment of a previously unavailable cognitive-semiotic capacity: condition (a) in the definition serves as the foundation, and the capacities listed under (b), (c) and the positive versions of (d) and (e), i.e. *with* conventionality/normativity (+d) and *with* semiotic systematicity (+e), constitute progressively "higher" stages in human socio-cognitive evolution. It is not a classical stage model in the spirit of Piaget where each consecutive stage brings with it total reorganization[16] but a "layered model" (Stern, 2000 [1985]) where earlier capacities continue to co-exist with newer ones, which may subsume but not abolish the earlier ones (cf. Section 2.2.5).

Condition (a) states that a special form of active perception (Berthoz, 2000, Section 2.2.1), in which (dynamic) aspects of the environment – especially the actions of con-specifics – are mapped onto one's own bodily actions and sensations, is a necessary condition for bodily mimesis. This gives rise to "shared representations" between self and other (Decety & Sommerville, 2003), or "self-other matching" (Barresi & Moore, 2008). However, this is not sufficient for true mimesis. If only condition (a) is fulfilled, as appears to be the case in neonatal imitation, the individual is capable of *proto-mimesis* but not bodily mimesis proper. Condition (b) states that a mimetic act needs to be both *volitional* and *representational*, as in Donald's original definition, and explicates the notion of representation in line with Piaget's double criterion of *both* differentiation *and* correspondence between "signifier" and "signified" from the subject's point of view (Piaget, 1945; Sonesson, 2007), adding the requirement that the signifier is a bodily act. Piaget's example of an infant opening and closing her mouth to model the opening and closing of a matchbox would be an example of an *iconic* correspondence (i.e. based on similarity) between the act and the object of attention. Children's acts of pointing for themselves in order to help guide their attention (Bates, Camaioni, & Volterra, 1975) would qualify as *indexical* (and more specifically *deictic*) mimetic acts (cf. Section 2.1.3, Section 4.5.2). This constitutes *dyadic mimesis*, which presupposes a "sense of a core self" (Stern, 2000 [1985]) in which the body is felt to be "one's own" and under volitional control (agency). This on its side is essential for one's body to serve as a representation of something else (as in Piaget's matchbox example) and to carry out precise imitation of novel actions.

Condition (c) introduces the necessary *triadic* element in order to make bodily mimesis communicative: the representation or sign is intended to be recognized as such (i.e. understood) by an addressee, as well as the communicative intention itself (Grice, 1989). The three-part relationship between self-initiated mimetic gesture, its meaning and the receiver of the intended meaning is what justifies calling this "triadic" mimesis. An example of an iconic sign that fulfils all three conditions (a), (b) and (c) is the *miming* of eating by pretending to move a spoon to one's mouth (e.g. made behind a glass door) in order to communicate to a

[15] "Stand for" may be further explicated in terms of the double asymmetry between expression and content discussed in section 2.1

[16] If this received interpretation of Piaget's developmental model is indeed correct, since Piaget recognized that figurative, "preoperational" thinking subsists even in the later "operational" periods.

colleague that it is time for lunch. An example of a mimetic sign that combines iconicity (since the motion and direction of the hand resembles the intended direction of attention of the addressee) and indexicality is *declarative pointing* (cf. Section 2.5.2).

If an act fulfils not only (a)-(c), but also the positive version of condition (d), then it is not properly speaking mimetic, but rather *symbolic,* since by adding the property of conventionality/normativity, the iconic and/or indexical motivation – or "ground" in Peircian terms (Sonesson, 2006) – of the sign loses much of its function, allowing the relationship between expression and content to become increasingly *arbitrary*. On this account, arbitrariness and conventionality are related, but non-synonymous notions, thus accommodating the fact that up to 50% of the signs of signed languages can be said to be more or less iconic (Woll & Kyle, 2004), but nonetheless conventional (and thus symbolic). A system of communication that has properties (a), (b), (c) and (+d) can be called a "protolanguage". Finally, by introducing the positive version of (e), *semiotic systematicity*, involving hierarchical relations between composite and simple signs (corresponding to what is usually referred to as "compositionality"), and furthermore relations to other signs (Deacon, 1997), the border between protolanguage and language has been traversed.[17]

Table 5.1. The Mimesis Hierarchy, defined by the attainment of a novel social-cognitive capacity and manifested in particular skills. Each stage corresponds to a condition in the definition of bodily mimesis (see text), with the last two stages "violating" the respective last two conditions. The different stages are related to similar, though non-equivalent notions used in the literature.

Stage	**Novel capacity**	**Cognitive/communicative skills**	**Related notions**
proto-mimesis	exteroception-proprioception mapping	- emotional and attentional contagion - neonatal imitation - mutual gaze	- sense of an emergent self (Stern, 2000 [1985]) - primary intersubjectivity (Trevarthen, 1979)
dyadic mimesis	volition and representation	- imitation - (cognitive) empathy - shared attention - mirror self-recognition	- sense of a core self (Stern, 2000 [1985]) - secondary intersubjectivity (Trevarthen & Hubley, 1978) - (true) imitation (Tomasello, 1999)
triadic mimesis	communicative intentions	- declarative pointing - iconic gestures - (full) joint attention	- pantomime (Arbib, 2005) - gesture (Corballis, 2002)
post-mimesis 1	conventionality/ normativity	- one-word utterances - holophrases	- protolanguage (Bickerton, 2003) - protosign (Arbib, 2005)
post-mimesis 2	semiotic systematicity	- spoken or signed language	- symbolic reference (Deacon, 1997) - tertiary intersubjectivity (Halliday, 1993) - narrative self (Stern, 2000 [1985])

This outline of the model is summarized in Table 5.1, relating the five different stages to particular cognitive-semiotic skills that testify to the attainment of particular novel capacities, and some non-synonymous but related concepts proposed in the literature on social learning and development. The model developed goes beyond Donald's model in several ways: (a) it suggests that a (typically) kinaesthesia-vision mapping serves as a precursor to bodily mimesis. (b) It distinguishes two different types of bodily mimesis: *dyadic*, which is not intentionally communicative, and *triadic*, which is – and implying a temporal precedence of the first over the latter in evolution and development. Furthermore, (c) within dyadic mimesis, in line with what was said in the previous section and in Section 2.2.4 we need

[17] It is possible that (+d) and (+e) always come in tandem since there are no existing languages which have only one but not the other, and Deacon has argued that this may be necessarily so due to "semiotic constraints" (Deacon, 2003). However, there are conceptual and empirical reasons for separating the two (Sonesson, 2007; Zlatev in press).

furthermore to distinguish between imitation, in which one's action stands in a token-to-type relation to what is imitated but is not "about" it, and (private) pantomime, in which the sign-function is realized, albeit not for the sake of someone else. (d) It emphasizes the aspect of conventionality as the crucial distinction between mimesis and language. Finally (e) the model implies a mutual dependency (and co-evolution) between the different levels, and levels of intersubjectivity (Zlatev 2008).

5.2.2. The mimesis hierarchy and evidence from SEDSU

In a manner similar to that done in previous research (e.g. Zlatev, Persson, & Gärdenfors, 2005; Zlatev, 2008, in press) we here relate the MH model to empirical evidence, though this time predominantly from the SEDSU project. Since the different levels/stages of the model apply to all Four Levels of Meaning: perception, imitation, sign function and convention (Section 3), results related to these summarized in separate sub-sections in Section 4 may be combined. On the other hand, we will be necessarily selective, since not all of the results can be brought to bear on the model.

5.2.2.1 Proto-mimesis

Three capacities which can be grouped under the heading proto-mimesis (i.e. involving self-other matching) were shown to be present in all the great apes: (a) recognition of other's imitation (Section 4.3.2) (b) mutual gaze (Section 4.4.3), though the rates of the second differed between the individual ape dyads studied, and even more so between these and the human dyads, and (c) synchronous perceptual intersubjectivity (Section 4.4.4).

On the other hand, chimpanzees were the only species shown to engage in yawning contagion (4.3.1), similarly to human beings (so far attested for children older than 4 years of age, Anderson et al. 2005). This difference between the species can be correlated with the finding that only chimpanzee (and bonobo) mothers displayed tolerance toward food sharing with their infants (4.4.2).

On the other hand, baboons (at least the subjects studied) were not able to recognize biological motion in point-light displays (4.1.2), which could also be due to a lower capacity for proto-mimesis. Chimpanzees (and apes in general) were unfortunately not tested in the paradigm, but if this capacity is indeed based on an implicit self-other mapping and given the results listed above, the prediction from the present analysis is that they would indeed be able to do so.

These somewhat tentative findings support the hypothesis for differences even on the level of proto-mimesis between (1) monkeys, (2) apes in general and (3) chimpanzees and bonobos, implying that the capacity for self-other mapping has evolved in evolution, above all as an adaptation for intersubjectivity. In previous formulations of the MH model (Zlatev 2008), it was suggested that there are "no qualitative differences" between human and e.g. chimpanzee infants in terms of proto-mimesis and hence primary intersubjectivity, but as some of the results showed (4.4.3 and 4.4.4), there are at least phenotypical differences between us and even our nearest relatives in the animal kingdom even on the level of proto-mimesis. These may be "quantitative", but nevertheless constitute important enough biases to lead to different developmental outcomes.

Children with autism did not have any difficult in with recognizing biological motion in general, but their difficulty in interpreting PLDs of emotional ("frightened") and to some extent subjective ("itchy") states, could also possibly be explained as an impairment affecting bodily mimesis at its core - in a developmental rather than an evolutionary context (Barresi and Moore 2008; Zlatev in press).

5.2.2.2 Dyadic mimesis

The study of do-as-I-do imitation in the enculturated chimpanzee Alex (4.2.3), supported earlier findings (Custance et al. 1995), that non-goal oriented action imitation is indeed possible, but in the far outskirts of the chimpanzee "zone of proximal development". This was at first surprising, given other recent robust findings (e.g. Buttleman et al 2007) of even "rational" imitation in the species. However, the explanation offered in Section 4.2.3 in terms of a difficulty to "translate" between the perception and execution of actions is perfectly consistent with the MH model, implying a uniquely developed capacity for this in human beings. Other research has suggested that this capacity is supported by a uniquely developed premotor-parietal-temporal "mirror neuron system" in humans (Iacoboni 2005, cf. Zlatev in press). However, the various forms of goal/object-focused forms of imitation, in which chimpanzees (and other apes) have performed more successfully do not require this type of "precise imitation" (Donald 2001), in which the movement of one's own body in closely controlled in order to match that of a model.

At the same time, the relatively successful performance of Alex on this task was matched by the finding that he apparently also understood the content of videos and photographs representing the actions to be performed: the most convincing positive result of "pictorial mode" processing of pictures by a non-human in the SEDSU project. Thus, in a single subject the two kinds of dyadic (non-communicative) mimesis: imitation and sign function, were positively correlated. Since chimpanzees (and other apes) have also been shown to be capable of matching between mirror images and themselves, this confirms (a) the close relation between imitation and sign use argued for on conceptual grounds in Section 2 and in the previous sub-section, and (b) earlier proposals on the basis of the MH model (Zlatev, Persson and Gärdenfors 2005) that dyadic mimesis as a whole is not beyond the capacity of non-human apes. As with proto-mimesis, however, we would here wish to emphasize the "quantitative" *difference* between ape and human capacity in this respect, a difference that would continue to place apes on a different developmental trajectory compared to (typically developing) children with respect to higher levels of the mimesis hierarchy.

SEDSU did not contain experiments on imitation by children with autism, but it is well established that they are impaired in this respect as well, as would be predicted given a "mirror neuron" model explanation of the deficit (Williams et al 2001). The SEDSU findings that are most consistent with this were those which showed atypical processing of photographs of human faces, and matching between pictures and sentences (4.3.3). As stated there, it was not a question of perceiving the stimuli as pictures or not – it was clear that they did so - but that the manner in which they perceived them was atypical: "non-configurational" for the faces, and a higher degree and accuracy in matching the pitch contours of sentences with graphs representing this than typical controls. Other pictorial stimuli on the other hand (cartoon faces) were processed by children and adults with autism configurationally, i.e. in the same manner as the typically developing controls.

This suggested that it was neither (a) a general impairment in interpreting context, (b) impairment in cross-modality, (c) weak central coherence or (d) lack of configurational processing that was at the root of ASD, but a more specific impairment related to bodily mimesis. Furthermore, atypical processing (and acquisition) of language in ASD is likely to be not the *cause*, but rather the *result* of such an impairment. This hypothesis follows from the logic of the MH model, which identifies language with post-mimesis.

5.2.2.3 Triadic mimesis

The most convincing reports within the comparative psychology literature of spontaneous iconic (Tanner 2004) and deictic (e.g. Leavens et al. 2008) gestures by captive apes have not been able to demonstrate (understanding of) communicative intentions in these species. (Some even argue against such "mentalistic" notions as a matter of principle.) Studies of non-linguistic communicative sign use in the project were not performed with non-human

primates and we cannot contribute to this debate on empirical grounds. The two cross-cultural studies of semiotic development in children (4.5) concerned this topic, however.

The MH model would predict that children enter the stage of triadic mimesis around approximately 14 months, when "declarative pointing" showing third-order mentality ("I intend that you recognize my intention to communicate X") emerges. The compliance study concerned children's receptive skills of adults' directives at a much earlier age (6.5-12.5 months), and argued against the children acquiring any "idea of intentions" or a "a recursive representational awareness of intentions-for-intentions" during this period. However, this does not contradict the MH model, which for this age would predict only dyadic mimesis and "second-order mentality" on the part of the infant, i.e. (empathetic) understanding of their parents' feelings, actions and intentions, and a "sense of self" that is strong enough to allow a choice between compliance or *non*-compliance. Furthermore, as pointed out in 4.5.3, evidence for a drop in compliance was found by the researchers in the period between 6.5 and 9 months, suggesting a possible discontinuity in the processes governing infant compliance.

The actions of bodily communication (ABCs) study (4.5.2), on the other hand, showed a stage dominated by triadic mimesis during the first period in the data (18-20 months) of (predominantly) non-conventional deictic and iconic (mostly 1st person perspective) gestures. The rather abrupt transition around 20 months, with onset of productive language and a "convention insight" can be interpreted as a reflection of the transition from triadic mimesis to post mimesis.

5.2.2.3 Post-mimesis 1 and 2

The difference between the two stages of post-mimesis is one in terms of systematicity: while the first constitutes a holophrastic protolanguage, the second should involve flexible and open-ended combinations of signs: words *and* gestures. The ABCs study supported this distinction in a rather novel way: not focusing on the combinatory properties of language, but on the way gestures (with or without object manipulation) combined with it. What was most characteristic about the "20-month revolution" (cf. 4.5.3) was that this marked the high-point of stereotypical communicative acts in which word and gesture were combined quite rigidly e.g. pointing and saying "there" or head-shaking and saying "no". What followed, apart from a gradual increase in MLU, was a *decoupling* of speech and gesture, allowing with time their "flexible" combinations. Since toward the end of the period, this flexibility was pronounced and MLU was over 2 (i.e. the average length of utterance was over 2 words, a hallmark of systematic word combinations) this can be interpreted as a "transition" to post-mimesis 2. However, there was no clear evidence for a "jump" in development, similar to that around 20 months. In other words, the change from post-mimesis 1 to 2 was gradual, or alternatively post-mimesis 1 was itself an "unstable", transitional stage.

This may be an aspect in which the relative parallelism between ontogeny and phylogeny, implied by the MH model, breaks down. The literature on language evolution commonly recognizes a hypothetical stage of "proto-language" (though its interpretation differ), and post-mimesis 1 corresponds to this. However, in the context of development, with children being to some degree biologically adapted, and their environments culturally so, for the use of complex verbal language, there is no natural reason to expect anything like a stable holophrastic stage, punctuated by a "compositional insight".

5.2.3 Summary

Combining aspects of the developmental and evolutionary perspectives, Figure 5.2 presents the stages of the MH model as *processes* of typical human development, marked by *relative* discontinuities (within an overall continuous framework), corresponding to the three transitions into dyadic mimesis (volition and representation), triadic mimesis (communicative intentions) and post-mimesis (conventions). The figure outlines how *relatively* minor differences in capacity between children and apes in terms of proto-mimesis (self-other

matching) may lead into *major* ones in terms of dyadic mimesis (imitation and non-communicative signs) and therefore apes failing to enter into triadic mimesis altogether, given that they are not explicitly enculturated into sign use and (proto)language.

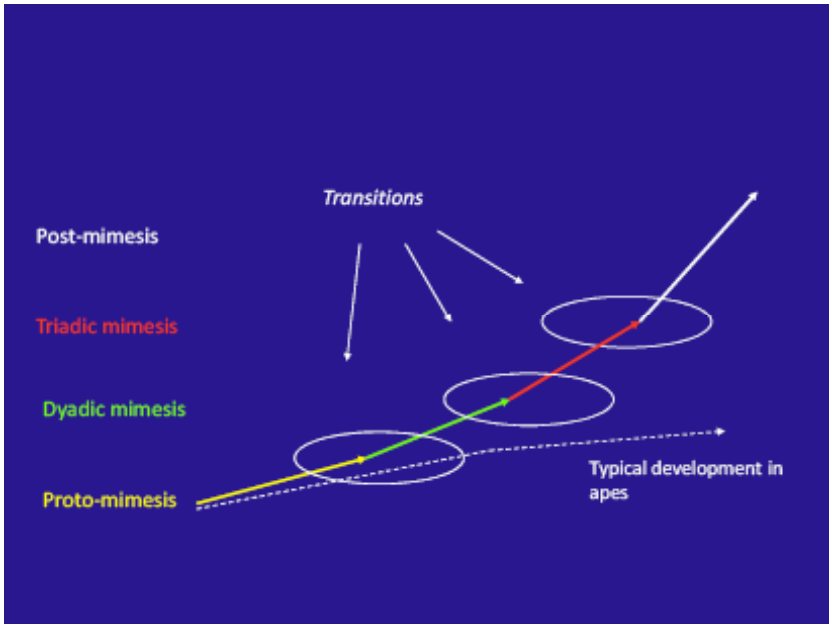


Figure 5.2. A schematic view of typical development in human beings and apes, with respect to the stages of the mimesis hierarchy model

The claim that sign use is presumed to arise from bodily mimesis (via imitation), also implies that *intersubjectivity*, and the forms of social interaction which both rely on it and promote its further development (e.g. mutual gaze, turn-taking, scaffolding) constitute essential, and possibly necessary pre-conditions for the emergence of sign use.

6 Conclusions and future research

In discussing the empirical findings for the SEDSU project in Section 4, we presented conclusions at the end of each subsection, and we do not wish to repeat this here again. But we have experienced that it is easy to "feel lost" given so many studies, methods, populations and results. Therefore we will make an effort here to present an overview of the major conclusions from the SEDSU project, "crossing" the Four Levels of Meaning (Perception, Imitation, Sign, and Convention, cf. Section 3) with the four types of comparative studies carried out (cross-species, cross-cultural, cross-populational and developmental, cf. Section 1).

Concerning the most basic level of meaning: *active perception* of the *Umwelt* one is embedded in, non-human primates seem to have a strong bias for local processing, *when they fail to see an ecologically meaningful object*, but rather patterns of light and colour. Human perception is in general both global (attending to the overall gestalt) and configural (attending the relations between the parts), but in certain conditions either of these can "break down", and perception revert to a more local strategy: as seen in the Himba case of visual illusions, or in the processing of faces in autism. One may wonder whether this is not due to similar factors as in the attested cases of local perception in non-human primates: when what is to be perceived is either "meaningless" or otherwise an "alien" (or threatening) part of the *Umwelt*. Differences in categorical perception concerning colour may be interpreted similarly. Baboons (for whom a categorical distinction was meaningless) lacked a categorical boundary between blue and green, Himba children and adults had a *different* boundary compared to British subjects, and these correlated with distinctions in both Himba language and environment (Lifeworld). Children with autism and controls had a similar boundary in terms of naming, but a different one in *sorting* (more closely based on perception). The conclusion we can draw so far is that meaning seems to determine perception, more than certain basic (pan-primate or pan-human) innate processing strategies. An effect of language, as a specific part of the Lifeworld, is possible, but should not be overestimated, given the evidence from autism, or the similar categorization of motion situations by Swedish and French speakers, possessing different linguistic resources.

With respect to *imitation*, the non-primate species most close to us, the chimpanzees, were only shown to have one of its precursors (contagious yawning), while all apes could recognize when they were themselves imitated. Do-as-I-do imitation was shown to be possible for a chimpanzee, but it turned out to be more difficult than what is commonly assumed. Experiments with imitation in autism were not performed, but difficulties are well-attested in the literature. Given that imitation was argued to be a "precursor" of the sign function for conceptual reasons, these findings support theories of semiotic evolution (e.g. Donald 1991, Arbib 2005, Zlatev in press) and autism (Williams et al. 2001), postulating a step-wise evolution (in autism: impairment) of a mirror neuron system.

The clear *use of signs* as defined in this report (Section 2) was only displayed by the non-human primate who had mastered a degree of imitation, the chimpanzee Alex, most conclusively by the fact that he could interpret photographs of incomplete actions correctly enough to be able to perform the actions represented. Since Alex was not in anyway language trained, this led to an important theoretical conclusion: that understanding pictures as signs does not require language. Many ingenious studies performed with capuchins (tokens, video-images, scale-models) showed advanced learning and memory for associations, but due to the nature of the tasks, we could not conclude that they demonstrate sign use. This was partly due to our definition symbols as conventional (rather than arbitrary) signs, and conventions require a capacity to entertain common knowledge, based on a collaborative relationship between the sign users.

Conventions were not attested, (or even tested) in non-humans, but basic behaviours requiring some degree of intersubjectivity such as food sharing were shown in apes and bonobos, to a greater extent than in other apes. On the other hand mutual gaze was found to be present to a much lower degree in ape than human mother-infant dyads, suggesting an evolutionary progression of primary intersubjectivity in the ape line. No general problem with learning conventional labels was found in autism, but the relation between perception (see above) and language was less close than in typical development, suggesting a partially different way of acquiring language. Unsurprisingly, radically different societies, living in different environments such as the Himba and the Amondawa had different linguistic and non-linguistic semiotic conventions compared to Western ones. In the case of Amondawa, Time was shown to be understood and expressed quite differently than in the West, but as with Himba colour concepts, whether language plays a causal role, or whether it "co-evolves" with culture remains a question for future research. In terms of development, the cross-cultural compliance study showed a pattern of gradual emergence, while the study of the ontogeny of acts of bodily communication (gestures), displayed a more qualitative shift around 20 months. In what was (somewhat provocatively) called a "convention insight" speech-gesture holophrases were clearly attained by Thai and Swedish children, giving rise to a gradual transition to more productive use of both speech and gesture.

The goal of formulating a novel, empirically informed and conceptually coherent theory of semiotic evolution and development was to some extent met by the Mimesis Hierarchy model, which presented an improvement (both empirically and conceptually) in comparison to Donald's stage model. In Section 5.2, we showed how it could account for results from non-primate perception (difficulties with point light displays), autism (given that configural processing of social stimuli is based on self-other matching), imitation and the development of conventions. Its nature as a "layered model" would also predict partial independence of the capacities on the different levels, thus allowing for perception to be relatively unaffected by language, unless the tasks rest on linguistic mediation. The model was formulated in terms of stages of semiotic evolution, but with adjustments and reservations, it could also be used to describe (and possibly eventually explain) the ontogeny of semiosis as well.

Given these conclusions, a natural direction for future research suggests itself, since the "level" of imitation in our conceptual framework which we take to be crucial for the attainment of the sign function, was the one least studied empirically (cf. the volume of the different sub-sections of Section 4). This is so because, to a considerable extent, it is only at the end of this research program that we have been able to distinguish clearly the many senses of what is ordinarily called imitation. We would therefore propose that future studies in semiotic evolution and development should focus precisely on the transition from imitation to sign use, in non-humans as in typical and atypical development. A related question concerns the precise relationship between intersubjectivity and imitation. Is one the foundation for the other, or are they so tightly connected that their development can only be described in terms of "co-evolution" (Zlatev 2008)?

In spite of all the experiments dedicated to it during the SEDSU project, picture interpretation still remains one of the mayor stumbling-blocks for any theory of semiotic evolution and development. Given the evidence from apes and children with some kind of language-related impairment, it seems that we can claim with some confidence that picture interpretation is not dependant on the mastering of language. Why then would it emerge, as suggested by the studies by DeLoache and Callaghan, only after language is attained? One possible explanation is that language is needed for the kind of tests to which language interpretation has been put. Another, which is not in contradiction to the first, is that what is tested is some kind of advanced picture understanding, which may perhaps have in common with language its being about the real world, not just about the fantasy world of pictures, familiar from children's picture books. New kinds of experiments will be needed to investigate these issues.

Finally, let us repeat the major goal of the SEDSU project, stated in Section 1: *To explain the evolutionary and ontogenetic development of sign use, clarifying its role for cognition.* It would be unfair to say that we have "explained" the evolution and development of sign use, but we have developed a conceptual apparatus, addressed empirical issues, and formulated empirical theories that can at least be said to have contributed to this sub-goal. Concerning the role of sign use for cognition, this is an even bigger topic, since many different forms of sign use need to be considered (e.g. the cognitive artefacts affecting our concept of Time), than we have had an opportunity to study systematically. But we can say that the SEDSU project has laid a ground for such future studies, since, as we have emphasized in this report, differences in sign use need to be distinguished from general Lifeworld differences, and the role of language needs to be disentangled from both.

7 References

7.1. List of Deliverables

Deliverable	Deliverable title
D1	Software system specification (WP2)
D2	Dissemination plan first version (WP1)
D3	Software for test procedure (WP3)
D4	Experimental and field linguistic protocols with preliminary results (WP4) – Portsmouth
D5	Digitised bi-cultural data corpus (WP5) – Lund
D6	First annual project review (WP0)
D7	Analysis of joint attention (WP6) – Lund
D8	Evolutionary analysis (WP2) – Marseille
D9	Analysis of mother-child interaction (WP6) – Lund
D10	Dissemination plan second version (WP1)
D11	Developmental analysis (WP2) – London
D12	Analysis of Cross-lingual results (WP2) – Goldsmiths
D13	Phylogeny of iconicity: an analysis (WP3) – Marseille
D14	Second annual project review (WP0)
D15	Cross-cultural study of compliance (WP6) –Portsmouth
D16	Summary: Perception and categorisation (WP7) – Goldsmiths
D17	Normal and impaired ontogeny of iconicity: an analysis (WP3) – Marseille
D18	Evolutionary analysis of landmark use (WP4) – Rome
D19	Summary: Intersubjectivty and conventions (WP7) – Lund
D20	Results from phylogenetic approach (WP5) – Leipzig
D21	Summary Iconicity and Pictures (WP7) – Marseille
D22	Results from ontogenetic approach (WP5) – Lund
D23	Analysis of spatial abilities in apes (WP4) – Rome
D24	Summary: Imitation and Mimesis (WP7) – Leipzig
D25	Field kit for investigating motion events and space based metaphor based on Amondawa and Thai (WP4) – Portsmouth
D26	Summary: Spatial Conceptualisation and Metaphor (WP7) – Portsmouth
D27	Overall Theoretical Summary (WP7) – Lund
D28	Final project review (WP0) – Goldsmiths

7.2 Other selected references

Addessi, E., Crescimbene, L. & Visalberghi, E. 2007. Do capuchin monkeys (Cebus apella) use tokens as symbols? *Proceedings of the Royal Society of London*, Series B, 274: 2579-2585.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioural and Brain Sciences , 28*, 105-167.

Arnheim, R. (1969). *Visual thinking.* Berkeley & Los Angeles, University of California Press.

Barbet, I. & Fagot, J. (2002). Perception of the corridor illusion by baboons. *Behavioural Brain Research*, *132*, 111-115.

Bard, K. A., Costall, M., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2005). Group differences in the mutual gaze of chimpanzees (Pan troglodytes). *Developmental Psychology , 41*, 616-624.

Barresi, J., & Moore, C. (in press). The neuroscience of social understanding. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen, *The Shared Mind: Perspectives on Intersubjectivity.* Amsterdam: Benjamins.

Bates, E. (1979) *The emergence of symbols*. Academic Press. New York.

Bentele, G. (1984) *Zeichen und Entwicklung. Vorüberlegungen zu einer genetischen Semiotik.* Tübingen: Narr.

Berthoz, A. (2000). *The Brain's Sense of Movement.* Cambrdige, Mass.: Harvard University Press.

Bråten, S. (2007). *On Being Moved: From Mirror Neurons to Empathy.* Amsterdam/Philadelphia: Benjamins.

Brinck, I. (2001). Attention and the evolution of intentional communication. *Pragmatics and Cognition,* 9/2: 255-272.

Brinck, I. (2004a). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly,* 3(4): 429-446.

Brinck, I. (2004b). Joint attention, triangulation and radical interpretation: A problem and its solution,

Dialectica, 58(2): 179-206.
Brinck, I. (in press). The role of intersubjectivity for the development of intentional communication. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen, *The Shared Mind: Perspectives on Intersubjectivity.* Amsterdam: Benjamins.
Bruner, J. (1998). Routes to reference. *Pragmatics & Cognition*, 6 (½): 209-227.
Byrne, R. W., & Tanner, J. E. (2006). Gestural imitation by a gorilla: Evidence and nature of the capacity. (T. Zentall, Ed.) *International Journal on Psychology and Psychological Therapy, Special Issue on Animal Learning and Cognition , 6* (2), 215-231.
Cabe, P. (1980) Picture perception in nonhuman subjects, in *The perception of pictures, Volume Il: Dürer's devices*. Hagen, Margaret, (ed.), 305-343. New York: Academic Press.
Callaghan, T. & Rankin, M. (2002) Emergence of graphic symbol functioning and the question of domain specificity. In *Child Development*, 73-2, 359-376.
Campos, J. J. & Steinberg, C. R. (1981). Perception, appraisal, and emotion: The onset of social referencing. In *Infant Social Cognition: Empirical and Theoretical Considerations*, M. E. Lamb and L. R. Sherrod (eds.), 273-314. Hillsdale, N.J.: Erlbaum.
Costall, A. (2007). Bringing the body back to life: James Gibson's ecology of embodied agency. In T. Z. Ziemke, & R. Frank, *Body, Language and Mind. Volume 1: Embodiment* (pp. 55-83). Berlin: Mouton de Gruyter.
Custance, D., Whiten, A., & Bard, K. (1995). Can young chimpanzees (Pan troglodytes) imitate arbitrary actions? Hayes & Hayes (1952) revisted. *Behaviour , 132*, 11-12.
Daddesio, T. C., (1995) *Of minds and symbols*. Mouton de Gruyter: Berlin & New York.
Deacon, T. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: Norton.
Deacon, T. (2003). Universal grammar and semiotic constraints. In Christiansen, Morton H, & Simon Kirby (eds)., *Language evolution,* 111-139. Oxford: Oxford University Press.
Decety, J., & Sommerville, J. (2003). Shared representations between self and others: A social cognitive neuroscience view. *Trends in Cognitive Science , 7*, 527-533.
Deely, J. (1982*). Introducing Semiotic. Its history and doctrine.* Bloomington: Indiana University Press.
Deloache, J. & Burns, N. (1994) Early understanding of the representational function of pictures. In *Cognition* 52, 83-110.
Deloache, J. & Miller, K. & Rosengren, K. (1997) The credible shrinking room. In *Psychological Science*, 8, 4, 308-313.
Deloache, J. (2000) Dual representation and young children's use of scale models. In *Child Development*, 71, 2, 329-338.
Deloache, J. (2004) Becoming symbol-minded. In *Trends in cognitive sciences*, 8, 2, 66-70.
Deregowski, J.B (1976) On seeing a picture for the first time, in *Leonardo* 9:1, 19-23.
Deregowski, J.B (2000). Pictorial perception: Individual and group differences within the human species. In Fagot, J. (Ed) *Picture Perception in Animals* (pp. 397-429). London:Psychology Press
Deregowski, J.B. (1973) Illusion and culture, in *Illusion in Nature and Art*. Gregory, R. L., & Gombrich, E. H., (eds.), 160-191. London Duckworth & Co, Ltd. Third impression 1980.
Donald, M. (1991) *Origins of the Modern Mind*. *Three Stages in the Evolution of Culture and Cognition,* Cambridge, Mass.: Harvard University Press.
Donald, M. (2001). *A Mind So Rare. The Evolution of Human Consciousness.* New York: Norton.
Emory, N. J., Lorincz, E. N., Perret, D. I., Oram, M. W. & Baker, C. I. (1997). Gaze Following and Joint Attention in Rhesus Monkeys (Macaca mulatta). *Journal of Comparative Psychology*, 111(3), 286-293.
Emory, N.J. (2000). The eyes have it. *Neurosci Biobehav Revue*, 24, 581-604.
Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford: Oxford University Press.
Gallagher, S. and Marcel, A. J. (1999). The self in contextualized action. *Journal of Consciousness Studies*, 6(4): 4–30.
Gardner, H. & Wolf, D. (1983) Waves and streams of symbolization: notes on the development of symbolic capacities in young children, in Sloboda, John, & Rogers, Don, eds., *The Acquisition of Symbolic Skills* ; pp. 19-42.
Gibson, J. (1966): *The senses considered as perceptual systems.* Boston: Houghton Mifflin Co.
Gibson, J. (1978). The ecological approach to visual perception of pictures, *Leonardo,* 4:2: 227-235.
Gibson, J. (1980). A prefatory essay on the perception of surfaces versus the perception of markings on a surface, in Hagen, Margaret, ed., *The perception of pictures, Volume I: Alberti's window*. xi-xvii. New York: Academic Press.
Gibson, J. (1982). *Reasons for realism.* Reed, Edward, & Jones, Rebecca, (eds). Hillsdale New Jersey: Lawrence Eearlbaum Ass.
Goldin-Meadow, S., & Butcher, C. (2003). Pointing towards two-word speech in young children. In S. Kita, *Pointing: Where Language, Culture and Cognition Meet* (pp. 85-107). Mahwah, NJ: Erlbaum.
Gómez, J-C (1998). Some thoughts about the evolution of LADS, with special reference to TOM and SAM. In P. Carruthers and J. Boucher (eds), *Language and Thought*. Cambridge: Cambridge University Press, 76-93.
Hobson, P. (2004). *The Cradle of Thought: Exploring the Origins of Thinking.* London: Pan.

Hurley, S., & Chater, N. (2005). *Perspectives on Imitation: From Neuroscience to Social Science. Vol 1: Mechanisms of Imitation and Imitation in Animals.* Cambridge, Mass.: The MIT Press.
Husserl, E. (1977 [1925]). *Phenomenological Psychology.* The Hague: Nijhoff.
Iacoboni, M. (2005). Understanding others: Imitation, language and empathy. In S. Hurley, & N. Chater, *Perspectives on Imitation: From Neuroscience to Social Science. Vol 1. Mechanisms of Imitation and Imitation in Animals* (pp. 77-100). Cambridge, Mass.: The MIT Press.
Itakura, S. (1994) Recognition of line-drawing representations by chimpanzees (Pan troglodytes). In *Journal of General Psychology*, 121, 189-197.
Itkonen, E. (1978). *Grammatical Theory and Metascience: A Critical Inquiry into the Philosophical and Methodological Foundations of "Autonomous" Linguistics* . Amsterdam: Benjamins.
Itkonen, E. (2003). *What is Language? A Study in the Philosophy of Linguistics*. Turku: Turku University Press.
Itkonen, E. (2008). The central role of normativity for language and linguistics. In J. Zlatev, T. Racine, E. Sinha, & E. Itkonen, *The Shared Mind: Perspectives on Intersubjectivity.* Amsterdam/Philadelphia: Benjamins.
Jones, R. & Hagen, M. (1980) A perspective on cross-cultural picture perception, in *The perception of pictures, Volume II: Dürer's devices*. Hagen, Margaret, (ed.), 193-226. New York: Academic Press
Kaminski, J., Tiedel, J., Call, J., & Tomasello, M. (2005). Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object choice task. *Animal Behaviour*, 69, 11-18.
Kennedy, J. M., (1974): *A psychology of picture perception.* San Francisco: Jossery-Bas, Inc.
Kuhlmeier, V. & Boysen, S. & Mukobi, K. (1999) Scale-model comprehension by Chimpanzee, in *Journal of comparative psychology,* 113, 4, 396-402.
Kuhlmeier, V. & Boysen, S. (2001) The effect of response contingencies on sclae model task performance by chimanzee, in *Journal of comparative psychology*, 115, 3, 300-306
Kuhlmeier, V. & Boysen, S. (2002) Chimpanzee recocognize spatial and object correspondeces between a scale model and its referent, in *Psychological Science*, 13, 1, 60-63.
Leavens, D. A., & Hopkins, W. D. (1999). The whole hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology , 113* (4), pp. 417-425.
Lewis, D. K. (1969). *Convention: A Philosophical Studie.* Cambridge, Mass: Harvard University Press.
Lotman, J.M & Uspenskij, B.A., Ivanov, V.V., Toporov, V.N., & Pjatigorskij, A.M., (1975). *Thesis on the semiotic study of culture.* Lisse: The Peter de Ridder Press.
Mandler, J. (2004) *The foundations of mind.* Oxford: Oxford University Press.
Meltzoff, A., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development , 54*, pp. 702-709.
Merleau-Ponty, M. (1962 [1945]). *Phenomenology of Perception.* London: Routledge.
Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2004). Imitation in neonatal chimpanzees (Pan troglodytes). *Developmental Science , 7/4*, 437-442.
Nelson, K. (1996). *Language in Cognitive Development: The Emergence of the Mediated Mind.* Cambridge: Cambridge University Press.
Peirce, C. S. (1931-58). *Collected Papers I-VIII.* Hartshorne, C, Weiss, P, & Burks, A, (eds.). Cambridge, MA: Harvard University Press.
Peirce, C. S. (1998). *The Essential Peirce, volume II.* Ed. By the Peirce Edition Project. Bloomington and Indianapolis: Indiana University Press.
Persson, T. (2008) *Pictorial primates*. Lund: Lund University.
Piaget, J. (1945). *La formation du symbole chez l'enfant.* Neuchatel: Delachaux & Niestlé. Third edition 1967.
Piaget, J. (1967). *La psychologie de l'intelligence.* Paris: Armand Colin.
Piaget, J. (1970). *Epistémologie des sciences de l'homme*. Paris: Gallimard.
Pika, S., Liebal, K., & Tomasello, M. (2005). Gestural communication in subadult bonobos (Pan paniscus): Repertoire and use. *American Journal of Primatology , 65*, 39-61.
Sonesson, G. (1989). *Pictorial concepts. Inquiries into the semiotic heritage and its relevance for the analysis of the visual world*. Lund: Aris/Lund University Press.
Sonesson, G. (1992). The semiotic function and the genesis of pictorial meaning. In Tarasti, Eero (ed.), *Center/Periphery in representations and institutions. Proceedings from the Conference of The International Semiotics Institute, Imatra, Finland, July 16-21, 1990.*, 211-156. Imatra: Acta Semiotica Fennica.
Sonesson, G. (1993). Pictorial semiotics, Gestalt psychology, and the ecology of perception. In *Semiotica* 99: 3/4, 319-399.
Sonesson, G. (1994). Prolegomena to the semiotic analysis of prehistoric visual displays. In *Semiotica,* 100: 3/, 267-332. 47-64.
Sonesson, G. (1995). On pictorality. The impact of the perceptual model in the development of visual semiotics. In Sebeok, Th., & Umiker-Sebeok, J. (eds), *.The semiotic web 1992/93: Advances in visual semiotics,* 67-108. Mouton de Gruyter, Berlin & New York.
Sonesson, G. (1996). An essay concerning images. From rhetoric to semiotics by way of ecological physics. In *Semiotica,* 109-1/2, 41-140.
Sonesson, G. (1998). /entries/, in *Encyclopaedia of Semiotics*. Bouissac, Paul, ed., New York & London: Oxford University Press.

Sonesson, G. (2000). Iconicity in the ecology of semiosis, In Troels Deng Johansson, Martin Skov & Berit Brogaard (eds.), *Iconicit y - A Fundamental Problem in Semioti,* 59-80. Aarhus: NSU Press.
Sonesson, G. (2001a). From Semiosis to Ecology. On the theory of iconicity and its consequences for the ontology of the Lifeworld. In Quinn, Andrew W., (ed.), Cultural cognition and space Cognition/Cognition culturelle et cognition spatiale. *VISIO*, 6: 2-3: 85-110.
Sonesson, G. (2001b). De l'iconicité de l'image à l'iconicité des gestes, *Actes du congrès ORAGE 2001 ORAlité et Gestualité, Aix-en-Provence, Juin 18-22, 2001.* Paris : L'Harmattan.
Sonesson, G. (2003a). The Symbolic Species revisited. Considerations on the semiotic turn in cognitive science and biology *SGBWP3. Working Papers of the Project "Language, gesture, and pictures from the point of view of semiotic development",* Lund University, Lund.
Sonesson, G. (2003b). Why the mirror is a sign – and why the television picture is no mirror. Two episodes in the critique of the iconicity critique. In *Proceedings of the 5th International Interdisciplinary Symposium of the Austrian Society for Semiotic Studies "Picture Language - Visualization – Diagrammatics: Iconicity/Das 5. Wiener Symposium 'Bildsprache - Visualisierung - Diagrammati": Ikonizität , November 2002*, published as *S. European Journal for semiotic studies,* volume 15:2-4, 2003, pp. 217-232.
Sonesson, G. (2007) The signs of the body and the body of signs. From ecology to semiosis in the theory of embodiment. In Zlatev, Jordan, Ziemke, Tom, & Frank, Roslyn, (eds.), *Body, language, and mind*. Berlin & New York: Mouton de Gruyter.
Sonesson, G. (Forthcoming a). From Iconicity to Pictorality. A view from ecological semiotics. To appear in VISIO: thematic issue: *Iconicity revisited/Iconicité revisité*, Sonesson, Göran (ed).
Stern, D. N. (1985). *The Interpersonal World of the Infant.* New York, NY: Basic Books.
Tanner, J. E., Patterson, F. G., & Byrne, R. W. (2006). The development of spontaneous gestures in zoo-living gorillas and sign-taught gorillas: From action and location to object representation. *Journal of Developmental Processes , 1*, 69-102.
Tomasello, M. (1998). Reference: Intending that Others Jointly Attend. *Pragmatics & Cognition*, 6(1/2), pp. 229-243.
Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.
Tomasello, M., Carpenter, Malinda, Call, Joseph, Behne, Tanya, & Moll, Henrike, (In press). Understanding and sharing intentions: the origins of cultural cognition. In *Behavioral and Brain Sciences,* in press 2004.
Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa, *Before Speech* (pp. 321-347). Cambridge: Cambridge University Press.
Trevarthen, C. and Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding, and acts of meaning in the first year. In *Action, Gesture, and Symbol: The Emergence of Language*, A. Lock (ed.), 183-229. New York: Academic Press.
Vygotsky, L. S. (1978). *Mind in Society: The Development of Higer Psychological Processes* . Cambridge, Mass.: Harvard University Press.
White, R. (2000). *Prehistoric art.* New York: Harry N. Abrahams, Inc.
Whiten, A., Horner, V., Litchfield, C. A., & Marshall-Pescini, S. (2004). How do apes ape? *Learning and Behavior , 32/1*, 36-52.
Williams, J. H., Whiten, A., Suddendorf, T., & Perret, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioural Reviews , 25*, 287-295.
Winner, Ellen (1982) *Invented Worlds*. Cambridge: Harward University Press
Zlatev, J. (2002). Mimesis: The "missing link" between signals and symbols in phylogeny and ontogeny. In A. Pajunen (ed.), *Mimesis, Sign and the Evolution of Language* (pp. 93-122). Turku: University of Turku Press.
Zlatev, J. (2003). Meaning = Life (+ Culture). An outline of a unified biocultural theory of meaning. *Evolution of communication 4:2*, 253-296.
Zlatev, J. (2005). What's in a schema? Bodily mimesis and the grounding of language. In B. Hampe, *From Perception to Meaning: Image Schemas in Cognitive Linguistics* (pp. 313-343). Berlin: Mouton de Gruyter.
Zlatev, J. (2007). Embodiment, language and mimesis. In T. Ziemke, J. Zlatev, & R. Frank, *Body, Language and Mind. Vol 1: Embodiment.* Berlin: Mouton de Gruyter.
Zlatev, J. (2008). The co-evolution of intersubjectivty and bodily mimesis. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen, *The Shared Mind: Perspectives on Intersubjectivity.* Amsterdam: Benjamins.
Zlatev, J. (in press). From protomimesis to language: evidence from primatology and social neuroscience. *Journal of Physiology, Paris* .
Zlatev, J., & SEDSU-Project. (2006). Stages in the Evolution and Development of Sign Use (SEDSU). In A. Cangelosi, A. D. Smith, & K. Smith, *The Evolution of Language: Proceedings of the 6th International Conference (EVOLANG6)* (pp. 379-388). New Jersey: World Scientific.
Zlatev, J., Blomberg, J., & David, C. (in press). Translocation, language and the categorization of experience. In P. Chilton, & V. Evans, *Space in Language and Cognition: The State of the Art.* London: Equinox.
Zlatev, J. Brinck, I and Andrén, M (in press) Stages in the development of perceptual intersubjectivity. In Enacting Intersubjectivity, F. Morganti , A Carassa and G. Riva (eds.) IOS Press.