

SEDSU Deliverable 7

An applied analysis of attentional intersubjectivity

Ingar Brinck, Jordan Zlatev & Mats André

Lund University

1. Introduction

The goal of the present deliverable is to provide *a developmental analysis of attentional intersubjectivity*, which, as we show below, is a more inclusive notion than the more commonly used term ‘joint attention’ (e.g. Moore & Dunham 1995). The use of the term ‘joint attention’ is not consistent in the literature, sometimes referring to the general phenomenon when two or more subjects attend to the same target (e.g. Butterworth 2003), sometimes to more reciprocal situations in which the subjects also are aware of attending to the same target (e.g. Tomasello 1999). Most often solely *visual* attention has been described, but implicitly the descriptions have been thought to generalize to other modalities. The concepts introduced in this deliverable constitute an attempt to construct a coherent framework that will allow for distinguishing and comparing the range of behaviours that in the literature have been addressed as ‘joint attention’ behaviours.

By *attentional intersubjectivity* we refer to the general case when *two or more subjects simultaneously focus their attention on the same target*. Attentional intersubjectivity will be further divided into *types*, according to which behaviours that are typically associated with attentional intersubjectivity occur during the interaction, and in which combinations. The result is that the over-all behaviour of the subjects during different types of attentional intersubjectivity differs. Our contentions are that:

- a) the types of attentional intersubjectivity identified in this report build on each other cumulatively and constitute different levels, and
- b) these levels correspond to *evolutionary* and *developmental* stages.¹

The analysis that we offer in this report is therefore of direct relevance for the SEDSU project (Zlatev et al. 2006).

While attentional intersubjectivity involves several perceptual modalities (at least vision, hearing, and touch), for practical reasons, this study primarily concerns the *visual* modality. Our analysis is, however, intended also to be applicable to these other modalities.

The analysis builds on previous work of the authors (Brinck 2001, 2003, 2004, in press a; in press b; Brinck & Gärdenfors 2003; Zlatev, Persson & Gärdenfors 2005; Zlatev in press), and provides a synthesis between *asymmetric* and *symmetric* attentional intersubjectivity – for definitions, see Section 2.

¹ In saying that the types are *cumulative* we mean that there is a progress by successive stages where each type is causally dependent on the type preceding it, and, furthermore, has increased in complexity as compared to previous types.

The more specific goal of this deliverable is to provide *operational definitions* of the different types and levels of visual intersubjectivity, which can be applied to various empirical data, in particular, the human and ape dyad corpora being collected within the SEDSU project: Deliverable 5 (*Digitized bi-cultural data corpus*), Deliverable 9 (*Analysis of mother-child interaction*) and Deliverable 15 (*Cross-cultural study of compliance*).² In this way, the theoretical model of attentional intersubjectivity is to be attested. The first steps to such an assessment are described in Section 4, where we apply the operationalized model to three sets of data:

- (1) interactions between adult and 4 infant chimpanzees and bonobos (from MPI Leipzig)
- (2) interactions between adults and 6 infants at approximately 18 months (from Deliverable 5)
- (3) interactions between adults and 2 infants of approximately 12 months (extra data for this deliverable)

Prior to this, in Section 2, we describe our model in fairly general, but still empirically attestable, terms, and in Section 3 we further operationalize the definitions to an extent that permits coding of video data of adult-infant interactions and testing of empirical hypotheses, such as those formulated in the pilot study described in Section 4. In Section 5 we conclude and suggest further applications of our model within the SEDSU project and elsewhere.

2. Types and levels of attentional intersubjectivity

2.1 Definitions

As stated in Section 1, **attentional intersubjectivity** (henceforth AIS) occurs when *two or more subjects simultaneously focus their attention on the same target*. The target can be classified as either:

- an **object**
- a **spatial location** (e.g., a certain place to go to)
- a **direction** (e.g., a way in which to go)

Although it may be difficult to distinguish the kind of targets on particular occasions, conceptually and ontologically the three are distinct (Jackendoff 1990).

The term ‘object’ will be understood in a wide sense to refer to an individual or entity that occupies a position in space-time, like a toy or a person. Events involving objects, such as the appearance of a new toy during on-going interaction between the subjects, or either of the subjects’ producing an action with an object, will similarly be categorised as ‘objects’ (though one can, of course, distinguish between objects in a more narrow sense and events).

Against treating directions as targets, it might be argued that a direction cannot be a target, because it does not specify a region in space, but merely a vector. We will nevertheless classify this case as one of attentional intersubjectivity, as the range of behaviours during the interaction will be similar whether or not the target is an object, location, or a direction in space. Having aligned their gaze, the subjects will not necessarily focus their attention on a

² There are several reasons why operational definitions in terms of objectively verifiable behavioural criteria *prima facie* have an advantage over functional definitions. See the discussion in Brinck (in press a), and also Bard (1992) and Leavens, Russell, & Hopkins (2005).

common, specific spatiotemporal region, as on an object or a spatial location, but sometimes will merely be looking in the same spatial direction. For instance, this will happen when a sound attracts the attention, and the source of the sound is not visible in the direction from which the sound was heard, as well as when only one of the subjects is able to detect the object.

Individual **episodes** of intersubjectivity will be individuated in terms of the concept of a target. A **target** is the object of undivided attention of one or both of the subjects. We will take an *AIS episode to have one, and not more than one, target*. The introduction of a new target for either of the subjects marks the end of the AIS episode.³

We claim that there are two major types of attentional intersubjectivity:

- **Symmetric:** when the target (T) has already been noticed by *both* (or more) subjects (Zlatev, Persson and Gärdenfors 2005; Zlatev in press). In discussing symmetric AIS, we will simply refer to the participants as **subjects**.
- **Asymmetric:** when the target (T) initially is noticed by *only one* of the subjects, and subsequently the other subject aligns his or her attention with the first subject's attention (Brinck 2001, 2003, 2004; Brinck & Gärdenfors 2003). We will call the first subject who has initially noticed the target the **sender** and the one who will focus her attention on the target as a consequence of the sender's behaviour the **receiver**.⁴

The relationship between the two types merits further research, but this is not our present focus. Our empirical analysis showed that it is easier to operationalize the asymmetric type, which will be the topic of the pilot study reported in Section 4. Nevertheless, the two types have *parallel* levels, and we capture this parallelism in the presentation in Section 2.

Prior to describing the different levels of symmetric and asymmetric AIS, we provide the following general definitions of the central terms that appear in the descriptions.

- **Focused attention:** the sender's or receiver's prolonged attention towards a target
- **Attention-getting behaviour:** the sender's behaviour directed at the receiver, which serves to make the receiver re-orient her attention toward the sender
- **Attention-turning:** the sender's orienting his attention towards the receiver
- **Attention contact:** the sender's and receiver's focused attention on each other's attentional state, in the case of visual attention consisting in **mutual gaze**
- **Gaze alternation:** the sender's alternating his gaze between target and receiver, performed relative to the attentional status of the receiver
- **Referential behaviour:** the sender's behaviour directed at the target while focusing on the target, serving to make the receiver re-orient her attention toward the *target*, such as *intention movements* performed in the direction of the target; a special case of social referential behaviour is pointing
- **Pointing:** the extension of the hand (with or without the index finger outstretched) or the goal-directed movement of the head and/or some other body part towards the target – performed for the benefit of the receiver
- **Reaching:** the subject's outstretched arm(s) and hand(s) in the direction of the target with the hand and fingers being formed as to grasp the target as the target is approached and the grip being adjusted as the distance to the target decreases

³ See Section 3 for more precise definitions.

⁴ We use these terms for ease of reference and not in their information-theoretical senses. In order to avoid clumsy gender-neutral expression like "he or she" or "(s)he" we will refer to the sender as "he" and the receiver as "she". This also allows us to manage the tricky business of "gender equality".

It is important to note that referential behaviours are not necessarily *social*, i.e. performed for the benefit of the receiver. Reaching can inadvertently to the sender acquire a referential function for the receiver. Pointing is by definition social, while ‘true’ reaching is performed for the benefit of the sender himself. Although social reaching gestures sometimes look similar to (individual) reaching behaviour, these behaviours can be distinguished. We will understand pointing in a broad sense as including reaching movements in case (a) the sender is *not persistently* trying to decrease the distance between himself and the target, and (b) the reach and the grip of the hand are *not adjusted* so as to fit the target. If on a particular occasion conditions (a) and (b) are satisfied, it indicates that on that occasion we have ‘pointing’ rather than ‘reaching’ according to our definitions.

2.2 Three levels of Attentional Intersubjectivity (AIS)

Based on the definitions given above, we distinguish three general levels of AIS, with symmetric and asymmetric counterparts. The first level is **Synchronous Attentional Intersubjectivity (SAIS)**, and is achieved by the subjects’ *globally relating their actions in time and space*. They are *performing similar individual actions* relative to a single target in the same spatiotemporal context. Synchronous intersubjectivity is not a social behaviour, because the individual actions are not performed for the benefit of another subject. Although the global behaviour may benefit other subjects, each individual action is designed to reach a personal goal.

The second level is **Co-ordinated Attentional Intersubjectivity (CAIS)**, and is achieved by the subjects’ *adjusting their actions relative to a single target*. On this level the behaviour is calibrated in time and space with respect to those of the other subject. The subjects are taking different roles vis-à-vis the target, and their respective behaviours will not interfere with each other. This behaviour is **social**, because the individual actions are performed for the benefit of the other subject, and may in circumstances to be specified below in addition be **interactive**, i.e., performed so as to directly affect the other (Brinck in press b). On this basis we can distinguish between two sub-levels of CAIS: social (Level 2.1) and interactive (Level 2.2) – see below for explicit definitions.

The third level is **Reciprocal Attentional Intersubjectivity (RAIS)**, and is achieved by the subjects’ *mutually matching their actions* relative to a single target. On this level each action is individually adjusted in space and time to the actions of the other subject. Similarly to Level 2.2, the behaviour is interactive, but in contrast the interaction between the subjects is more complex on Level 3 than level 2.2. 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 complementary turn-taking occurring during dyadic engagement, and also during ritualised behaviours such as the give-take game.

Level 1: Synchronous Attentional Intersubjectivity (SAIS)

Symmetric SAIS

T has independently captured subject A’s and subject B’s attention and has caused both A and B to focus their attention on T. (See Figure 1) Example: T is food or a predator, or something else of a similar intrinsic organismic value (Zlatev 2003) for both A and B.

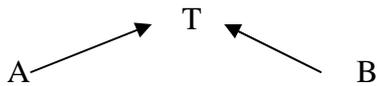


Figure 1. Symmetric Synchronous Attentional Intersubjectivity

Asymmetric SAIS

Characterized by the following stereotypical sequence (See Figure 2):

- (1) A’s attention is attracted by T, T captures A’s interest and makes A focus his attention on T.
- (2) B’s attention is attracted by A’s attention-focusing.
- (3) B follows A’s orientation to T, with the result that both focus their attention on T.

Example: *attentional contagion*, e.g. when a goat re-orient its attention, focusing on some food being shown behind the back of another goat. The first goat’s attention behaviour towards the food will make the second goat look towards the first goat, and then re-orient its attention to the food by following the direction of the first goat’s attentional state (cf. Kaminski et al. 2005).

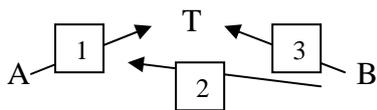


Figure 2. Asymmetric Synchronous Attentional Intersubjectivity

Level 2: Co-ordinated Attentional Intersubjectivity (CAIS)

Symmetric CAIS

T has already been noticed by A and B. In addition, A directs his or her attention to B’s attention-focusing on T, and B directs his or her attention to A’s attention-focusing on T. In contrast to Level 1, we have here *second-order attention* for both participants (both perceive that the other perceives T). See Figure 3. As example is the following situation of *social referencing*: A is an infant and B is an adult. T is of ambiguous value. By checking whether B is paying attention to T, and looking for indications of positive or negative reactions on the part of B, A can adjust his or her attitude to T. By monitoring A’s attention and attitude towards T, B can check if A is behaving appropriately towards T.

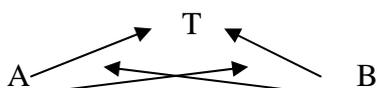


Figure 3. Symmetric Co-ordinated Attentional Intersubjectivity

Asymmetric CAIS

As mentioned at the beginning of this section, Asymmetric Co-ordinated Attentional Intersubjectivity can be divided into two sub-types, according to whether or not the infant is turning to the adult (and focusing his attention on the adult) during the interaction. *Social* behaviour is performed for the benefit of the receiver, while *interactive* behaviour is performed so as to directly affect the receiver.

In *Social CAIS*, the infant ostensibly attends to the object by, e.g., gazing at and behaving referentially towards it (sometimes while vocalising), thereby making his behaviour perceptually accessible to the adult, so that the adult can notice the infant's attending to the object. However, the infant does not engage with the adult by turning towards the adult to check whether the adult is attending to the infant himself.

In *Interactive CAIS*, the infant directs himself straight to the adult by turning his attention to the adult. The infant thereby draws the adult's attention to himself and his behaviour directed at the target. Attention-turning sometimes occurs together with attention-getting behaviour.

Level 2.1: Social CAIS

Characterized by the following stereotypical sequence of behaviours:

- (1) A focuses his attention (typically gaze) on T.
- (2) A *ostensively gazes at T and engages in referential behaviour towards T*.
- (3) B notices (2).
- (4) B follows A's gaze and referential behaviour and notices T.

Example: simpler forms of *imperative pointing*, when an infant points toward an object, without interacting directly with the adult. Note that the novel behaviour that distinguishes Level 2.1 from Level 1 is (2), the referential behaviour and manifest gaze (and vocalisation, if present).

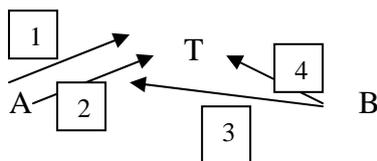


Figure 4. Asymmetric Social Co-ordinated Attentional Intersubjectivity

Level 2.2: Interactive CAIS

Characterized by the following stereotypical sequence (see Figure 5):

- (1) A is focusing his attention (typically gaze) on T.
- (2) A engages in attention-getting behaviour directed at B. (optional)
- (3) A *ostensively gazes at T and engages in referential behaviour towards T, turning his attention towards B*.
- (4) B's attention is attracted by either (2) or (3), then B focuses her attention on A.
- (5) B follows A's referential behaviour and/or gaze to T.

Example: typical cases of imperative pointing, in which a child “makes sure” that the adult is attending before or while performing the pointing gesture. Note that the crucial behaviour that distinguishes Level 2.2 from Level 2.1 is (3) in which the infant directs his attention towards the adult during the interaction. (2) also constitutes a difference, but since (2) is optional, it cannot be used to distinguish the two sub-types.

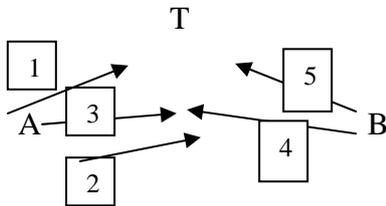


Figure 5. Asymmetric Interactive Co-ordinated Attentional Intersubjectivity

Level 3: Reciprocal Attentional Intersubjectivity (RAIS)

Symmetric RAIS

A not only attends to B’s attention to T and vice versa (as in symmetric SA, Level 2), but A attends to B’s attending to A’s attention, and vice versa: on this level, we have *third-order attention* (Zlatev in press). Example: A child and an adult play a game with hiding toys. The child sees the hidden toy, smiles, and then looks at the adult and sees that the adult sees that he has seen the toy. Both acknowledge this (verbally). See Figure 6, where only the third-order attention of the child (A) is shown.

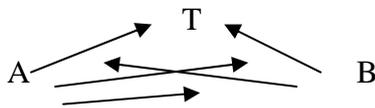


Figure 6. Symmetric Reciprocal Attentional Intersubjectivity: involving third-order attention (here shown only for A in the arrow at the bottom).

Asymmetric RAIS

Characterized by the following stereotypical sequence (see Figure 7):

- (1) A is focusing his attention (typically gaze) on T.
- (2) A engages in *attention-getting behaviour* relative to B. (optional)
- (3) B’s attention is attracted by (1) or (2), and B focuses her attention on A.
- (4) A and B exchange gazes, establishing *attention contact* (double arrow in Figure 7)
- (5) A ostensibly gazes at T and/or engages in referential behaviour towards T.
- (6) B follows A’s pointing gesture and/or gaze to T.

Example: typical cases of *declarative pointing*. Note that the novel behaviour, distinguishing Level 3 from Level 2, is (4), (visual) attention contact, i.e. mutual gaze.

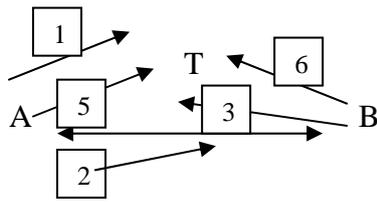


Figure 7. Reciprocal Asymmetric Attentional Intersubjectivity

2.3. Summary

In this section we have provided a level-based analysis of attentional intersubjectivity, where each consecutive level is of higher *complexity* than the previous one.

In the case of symmetric AIS, where the target has already been noticed by the participants of the interaction, this complexity can be defined as first-order attention (Level 1), second-order attention (Level 2) and third-order attention (Level 3).

In the type of AIS which we call asymmetric, due to the fact that initially only one of the subjects has noticed the target and the other does so due to the attentional behaviour of the first, the different levels are defined by sequences of behaviours. In this type of AIS the increased complexity is reflected by the fact that each higher level *subsumes* the previous ones, and also includes novel behaviours: On Level 1 the behaviour of the sender is not social (in sense of being performed for the benefit of the receiver) but individual. In contrast, on Level 2 (co-ordinated AIS) the sender engages in various forms of referential behaviour (the clearest case of which is pointing). On Level 2.1 the sender engages in ostensibly manifest, behaviours for the benefit of the receiver, but does not turn his attention to the receiver to check if his signalling has been perceived. In contrast, on Level 2.2, which we call interactive, such attention turning occurs. Finally, Level 3 adds mutual gaze, during which the subjects simultaneously attend to each other's attentional states, which in the visual modality results in mutual gaze.

The different types and levels of attentional intersubjectivity are distinguished on the basis of overt behaviours and their sequencing. The present model deals with observable behaviours only, aiming to identify key behaviours descriptively, yet without interpreting them in psychological or socio-cultural terms to avoid unwarranted assumptions (cf. Brinck in press a). Hence the controversial notion of *intention* has been avoided in the definitions. The model is deliberately constructed in such a way as to rather yield “misses” than “false hits”, or in short, to be conservative. This is mandatory in order to substantiate our claim in the introduction that the levels that we have identified are *evolutionary* and *developmental*, i.e., correspond to capacities in different species more or less closely related to us, and to different stages of human ontogeny. Thus the model is not ignoring issues connected to intentionality, the role of language and various cultural practices, but these are left out of the account as the result of a carefully chosen strategy.

The rationale behind our methodology is its purpose, i.e., to construct a global model of attentional intersubjectivity that in all its varieties can be equally applied to empirical video-corpus data of adult-infant interactions in different species, cultures and at different ages, such as the data made available by the SEDSU project.

3. Further operationalization of the model of attentional intersubjectivity

The definitions of the different types and levels of AIS formulated in Section 2 were formulated while observing the data in the Thai/Swedish video-corpus (Deliverable 5). As mentioned, these definitions are meant to be empirically attestable and applicable to both human and non-human subjects. However, in order to be able to use them as the basis for a *coding scheme* in that analysis of such data, they need to be even more concretely specified, which is the goal of the present section.

A general guiding principle has been to be *conservative*, i.e. to have operational definitions which preferably “underinterpret” than “overinterpret” the observational data, especially that from the 18-month old children. The reason for this is that these children have already made their entrance into language, and as known, language can *substitute* for many other forms of intersubjective behaviour, including mutual gaze and gesturing (Tomasello 1999; Butterworth 2003). While being in essence a form of “referential behaviour” we have decided explicitly not to code the children’s utterances as such, since that would place them on an uneven footing compared to the behaviour of pre-linguistic children and apes. Furthermore, especially with the 12-months old subjects, it is not easy to distinguish verbalization from vocalization. Hence, in the operational definitions offered below, we treat vocalization as a form of *attention getting* (cf. Section 2), but not as any of the other crucial behaviours.

The first specification compared to the definitions of asymmetric AIS in Section 2, is that we analyze only cases in which the sender is the *infant*⁵ (human or ape) and the receiver as the *interacting adult* (parent or some other individual).

Furthermore, as mentioned in the introduction, we intend the definitions given below only to apply to the asymmetric variety of AIS, the reason being that it is much easier to individuate the AIS episodes for this type, rather than for symmetric type. The operational definition for how this was done is the following:

- **What counts as an “AIS episode”?**

The beginning of a new **AIS episode** is marked by the introduction of a **new target**. A target is an object, location or direction that receives undivided attention from one or both of the subjects. Therefore, an AIS episode by definition includes one, and not more than one, target and the introduction of another new target defines the end of the previous episode and the start of a new one. What operationally counts as a new target is based on *observable contrasts* in the subjects’ behaviour while interacting, rather than being based on properties of the targets themselves. New targets are judged to occur in the analysis in the following situations:

1. There is a shift in the infant’s attention to a target that is altogether outside his earlier focus of the interaction. In most cases the target is an object, but it may also be a sound, such as a telephone signal being heard in the background.⁶

⁵ For the sake of simplicity, we also refer to the 18-month-old children as ‘infants’.

⁶ Episodes resulting from salient sounds tend to result in symmetric AIS episodes rather than asymmetric ones, which is an interesting difference between modalities to investigate further in the future. For the present analysis (and the study in described in Section 4), we analyze sound-targets only when they lead to asymmetric AIS episodes, i.e. when only the infant seems to notice (and comment) on them at first.

2. There is a shift in the infant's attention to a target which is more or less within the earlier focus of the interaction, but a re-orientation of attention is observable in both the infant's and the adult's behaviour, such as:
 - a. *An object is singled out by/in contrast to other targets in the context.* Example: The activity consists in playing with building blocks. Although both participants are already attending to the building blocks in general, the infant may focus his attention on a specific building block while picking it up and thereby introducing a new target. In addition to this, the adult also visibly redirects her attention to this specific object.
 - b. *A part of an object is singled out by/in contrast to the object as a whole.* Example: The activity consists in playing with a toy telephone. Although the toy telephone is already the focus of the interaction the infant may shift his attention to a specific *part* of the telephone such as the comparatively small area of the telephone such as the mouthpiece. In addition to this, the adult also visibly redirects her attention to this specific part of the telephone.
 - c. *An object is being moved to another location, which will constitute a new target of attention.* Example: The sender is holding a glass of milk, and the glass is the focus of the interaction. Then the sender puts it down outside the current visual field of the receiver, who will shift her attention to this new location.

Since behaviors are to be classified into different categories, there must be a visible contrast in behavior available to the analyst to distinguish between these categories. This is a rather trivial point, but it has some implications for what might count as an episode at a certain level or a new target:

1. We are primarily interested in classifying the behavior of the infant. This means that in the cases mentioned above in the discussion of new targets, where new targets are altogether outside the focus of the ongoing interaction, as in case (1) above, it is not of crucial importance whether the target is already within the visual field of the adult as long as the infant cannot see this. It may still be an asymmetric AIS episode from the viewpoint of the infant. However, in cases (2a)-(2c) above, it is of crucial importance whether there is also a slight adjustment in the attention of the adult, if the target is already within the visual field of the adult. Otherwise it is impossible to establish that these more subtle kind of new targets really are established as common to both sender and receiver.
2. Since it is not possible to distinguish the infant's attention to the target from attention turning in cases where the target is the adult herself, these targets/episodes are excluded from analysis. In other words, our analysis deals with triadic only, and not dyadic, engagements.

Given these qualifications, we can operationally define the different levels of asymmetric AIS as follows:

- **Level 1 AIS**

The infant is focusing his attention (visual or auditory) on a new target, and the adult follows the infant's attention to this new target. However, the infant neither produces any referential behaviour, nor turns his attention to the adult.

INFANT: ATTENTION TO NEW T

ADULT: FOLLOW ATTENTION(INFANT) TO T

- **Level 2.1 AIS**

The infant is focusing his attention on a new (visual or auditory) target and engages in referential behaviour, but does not turn his attention to the adult. The adult follows the infant's attention and/or referential behaviour to the target.

INFANT: ATTENTION TO NEW T

+ REFERENTIAL BEHAVIOUR TOWARDS T

ADULT: FOLLOW ATTENTION(INFANT) TO T

- **Level 2.2 AIS**

The infant is focusing his attention on a new (visual or auditory) target and engages in referential behaviour, and also turns his attention to the adult. The adult follows the infant's attention and/or referential behaviour to the target.

INFANT: ATTENTION TO NEW T

+ REFERENTIAL BEHAVIOUR TOWARDS T

+ TURN ATTENTION TO ADULT

ADULT: FOLLOW ATTENTION(INFANT)TO T

- **Level 3 AIS**

The infant is focusing his attention on a new (visual or auditory) target and engages in referential behaviour, and also turns his attention to the adult. Furthermore, the infant and adult engage in attention contact (mutual gaze) – *during the infant's turn*, i.e. prior to the adult's verbally commenting on the target. The adult follows the infant's attention and/or referential behaviour to the target.

INFANT: ATTENTION TO NEW T

+ REFERENTIAL BEHAVIOUR TOWARDS T

+ TURN ATTENTION TO ADULT

+ MUTUAL GAZE WITH ADULT IN OWN TURN

ADULT: MUTUAL GAZE WITH INFANT

+ FOLLOW ATTENTION(INFANT) TO T

4. Pilot study of AIS analysis

The definitions described in Section 3 were used in the following pilot study.

4.1. Data

We analysed the following sets of data, which were made available by our and our colleagues' research within the SEDSU project.

- 2 video-recordings of an infant bonobo (Luiza, age 10 months) and chimpanzee (Lobo, age 19 months) collected by Mathias Osvath at MPI Leipzig, appr. 60 minutes each. Additionally, 2 video recordings of Luiza (at the age of 13 months) and the chimpanzee Kara (age 7 months), 8:30 minutes each, recorded by Josep Call and his assistants.
- 6 video-recordings from the Thai/Swedish video-linked corpus (Zlatev, Andrén & Osathanonda, Deliverable 5, WP6), involving three Swedish children (BEL, TEA and HAR) and 3 Thai children (JOM, JAM and CHE), when these children were app. 18 months old.

Our initial intention was to use data from Deliverable 15, *Cross-cultural study of compliance*, managed by Vasu Reddy at Portsmouth to study AIS in pre-verbal children. But since we were pressed for time and could not access this data immediately, we chose to analyze:

- 2 video-recording of 2 Swedish children aged 12 months: ALI (recorded by Mats Andrén, 15 minutes) and TEA (recorded by Ulla Richthoff, 23 minutes).

4.2. Hypotheses

As obvious from the above description, with the exception of the 2 data points from TEA and the Luiza, the data was not longitudinal, and hence did not permit us to test any explicit developmental hypotheses. However, given that the data was, in broad terms, cross-sectional, and the different AIS levels are of increasing complexity, we could formulate three broad hypotheses to test whether these correspond to developmental and evolutionary stages:

H1. AIS episodes of Level 3 (Reciprocal) will be attested predominantly among the 18-month old children.

H2. AIS episodes of Level 2 (Co-ordinated) will be observed among the 18-month old children and the 12-month old children.

H3. AIS episodes of Level 1 (Synchronous) will be the only form of attentional intersubjectivity found in the ape data – perhaps with occasional instances of Level 2.

4.3. Analysis

All AIS episodes (according to the definitions in Section 3) were identified in the data, resulting in a total of 190 instances, divided by the different video-recordings ('data points') as shown in Table 1.

Table 1. Total number of AIS episodes per data point

Data point	Age (months)	Length of data point (minutes)	# AIS episodes
Kara (chimp)	7	8:30	2
Lobo (chimp)	19	60	1
Luiza (bobobo)	10	60	1
Luiza (bonobo)	13	8:30	1
TEA (Swedish)	12	23	26
ALI (Swedish)	12	15	24
TEA (Swedish)	18	15	26
HAR (Swedish)	18	15	18
BEL (Swedish)	18	15	33
JAM (Thai)	18	15	11
JOM (Thai)	18	15	23
CHE (Thai)	18	15	24

2.4. Results

As the general analysis in Table 1 shows, asymmetric attentional intersubjectivity (as operationalized) seems to be a human speciality – not in the sense that it is *unique* for *Homo sapiens*, but that it is much more frequent in human infant-adult interactions. In approximately 2 hours and 18 minutes of data, the apes engaged on only 5 asymmetric AIS episodes. In contrast, the 2 hours and 8 minutes of human data contained a total of 185 instances!

Furthermore, the hypotheses – which were formulated prior to any data analysis – where almost surprisingly well confirmed, as shown in Figure 8. Level 3 episodes did not occur only in the interactions of the 18-month old children, but they were proportionally more frequent than for the two 12-month old children. Level 2 episodes occurred in both the data of the two groups of children, but were altogether absent (along with Level 3 episodes) in the ape data. All 5 instances of asymmetric AIS episodes initiated by the infant apes were of Level 1.

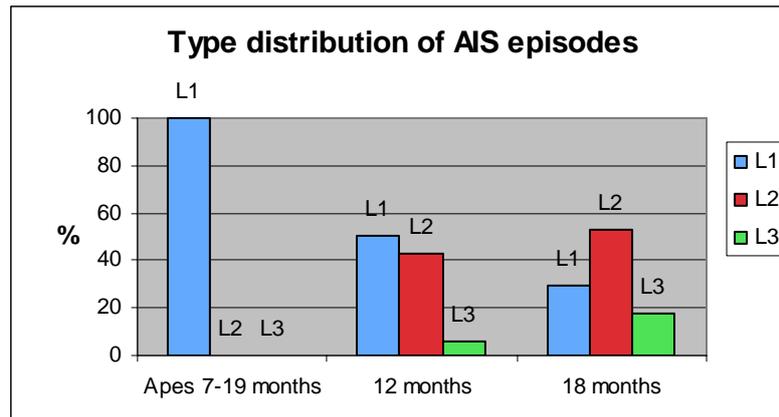


Figure 8. Percentages of asymmetric AIS episodes by type in the three sets of data. Total number of episodes: 5 for apes, 50 for 12-month old children, and 135 for 18-month old children

4.5. Discussion

The results of the pilot study give support to our proposal that the different levels of AIS in our model, at least of the asymmetric variety, correspond to evolutionary and developmental stages. The three ape infants studied (in the four data points) and their interacting adults engaged in a surprisingly low number of AIS episodes, and all of these were of the simplest type, Level 1 (synchronous), which are not social from the point of view of the infant. Even for adults, this type does not require anything more than *attentional contagion*, which is a form of perceptual cueing, spreading automatically by attention attraction (Brinck 2001, in press a).⁷ For instance, a movement of the sender occurring as a reflex response to some event in the vicinity of the subjects can inadvertently cause a behavioural co-ordination between sender and recipient, by incidentally attracting the attention of the recipient to the sender without engaging conscious awareness. Of course, this does not exclude that in subjects capable of higher-level processing than attentional contagion, other kinds of processing support the process.

The complete absence of any “higher” type of AIS episodes in the ape data, irrespective of the differences in the ages of the ape infants (7-19 months) supports the analysis of Level 1 AIS as being qualitatively different from the higher types and a corresponding difference between the species *Pan* and *Homo*. At the same time, we need not interpret this as a matter of *inability* of apes to engage in higher types, since we know from previous research that in captivity (adult) apes do engage in referential behaviour directed to human receivers (e.g. Leavens, Hopkins & Bard 1996; Leavens & Hopkins 1999). Nevertheless, the differences are so conspicuous, even in this pilot study, that they seem to reflect a qualitative difference in the nature of ape and human *social interactions*: human infants (and young children) engage in referential behaviour for the sake of a receiver on a regular basis, while ape infants do not. Furthermore, since the differences between the pre-verbal and “just-verbal” children with respect to this (as reflected in the minor differences in Level 2 in Figure 1 for the two groups) suggest that this is a feature of human social interactions that is *independent of and more basic than language*. In terms of the model proposed by Zlatev, Persson & Gärdenfors (2005) and Zlatev (in press), this difference can be interpreted as due to differences in dyadic

⁷ Neurally, attentional contagion appears to be supported by a specialised mechanism of the attention system in the brain (Brinck in press b; Chawarska et al. 2003; Driver et al. 1999). It exists on a subpersonal processing level that cannot be accessed or influenced by conscious awareness. The attention system immediately reacts to the perceivable re-orientation of body, head, or gaze, or all of these, of other subjects, and will cause the receiver of this signal to re-orient her own attention unless the behaviour is inhibited.

and – especially – triadic *bodily mimesis*. But other interpretations are, of course, possible (cf. Tomasello et al. 2005).

Probably the undeniable differences in intersubjective behaviour between apes and humans can be properly understood only in an ecological context. The fact that apes point in captivity, but not in the wild, shows that individual behaviour develops in dynamic interaction with the environment, and that it is vital to consider environmental factors when doing comparative research in view of drawing evolutionary conclusions (Leavens, Hopkins & Bard 2005). There is also evidence that chimpanzee infants, like human ones, have an innate predisposition for primary (dyadic) intersubjectivity, and that this can develop into secondary (triadic) intersubjectivity, e.g., social referencing (cf. Bard 2005; Tomanoga et al. 2004).

Because humans and apes parted lineage several million years ago, it is very likely that due to the increasing differences between their respective natural and social environments, their capacities for communication have developed along diverging paths. For instance, human intersubjective interaction takes place at a distance, while for apes it tends to occur while being in physical contact. This allows apes to convey information about their own and others' attentional states through other, more discrete and subtle media than by gaze and gesturing, e.g., by touch and sounds that are audible only for those who are positioned next to the sender. Humans can use gaze efficiently to communicate about distant objects; on the other hand, apes are very efficient in picking up information by quick glances. What seems to be global differences in behaviour may equally correspond to global differences in capacities as to global similarities (with the capacities *manifesting* themselves in separate ways). It is hard to draw any definite conclusions about the nature and basis of the divergent intersubjective behaviours in apes and humans at this stage.

However, the conclusion that human infants have a special knack for triadic mimesis seems correct. It is as well supported by the only piece of direct human developmental evidence in our data: While Level 1 episodes predominate over all the others for the child TEA at 12 months, occurring roughly once a minute in the interaction, the fall down to 0,4 per minute at 18 months, which Level 2 (both simple and asymmetric) AIS events rapidly increase (cf. Figure 9).

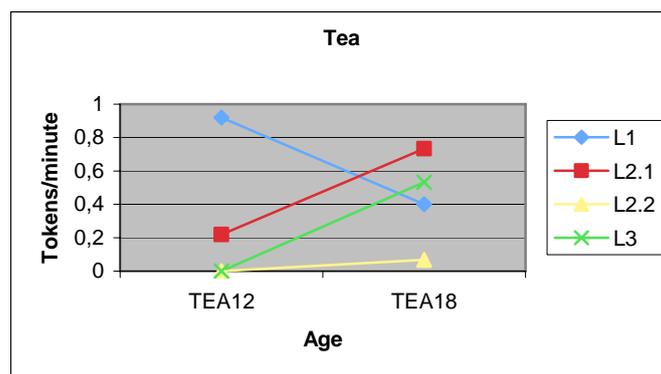


Figure 9. Number of AIS events per minute for TEA at 12 and 18 months.

The results concerning our first hypothesis (H1) regarding Level 3 (reciprocal) AIS dominating in the 18-month old group, the results were less clear cut. Indeed, there was a higher proportion of reciprocal AIS episodes in that group, but the 12-month old group had only 2 children, and there was considerable individual variation between the children in both

groups. Figure 10 shows the average number of AIS episodes per minute, divided by types, for ALI, aged 12 months, and HAR, aged 18 months, and as can be seen ALI has more Level 3 episodes, while HAR has mostly Level 1 episodes. On the other hand, HAR is the child with slowest linguistic development in the Swedish group (cf. Deliverable 5), and ALI seems (at least in the studied interaction) to be quite precocious. It is also interesting that the only Thai child with a similar AIS profile to HAR – JAM – is also characterized by slower linguistic development, compared to the other two Thai children: See Figure 11 where JAM is compared to JOM, who is both gesturally and linguistically a faster developer. Hence we tentatively conclude that Level 3 AIS does indeed correspond to a higher developmental level, though the correlations between AIS and language development need to be studied more close in the future.

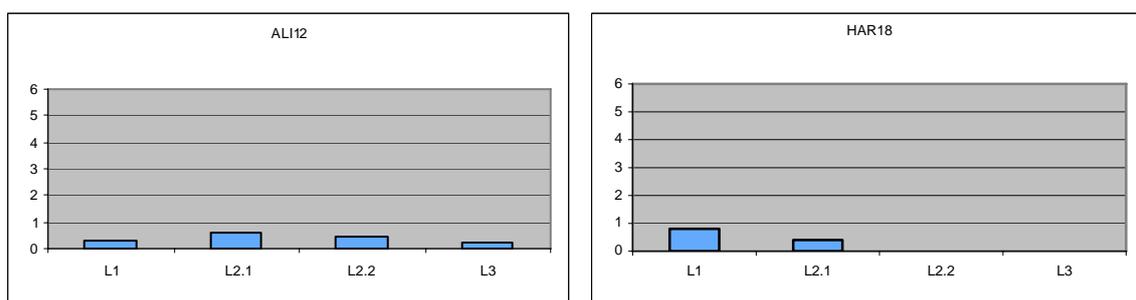


Figure 10. Number of AIS episodes per minute, divided by types for ALI (12 months) and HAR (18 months), both Swedish children.

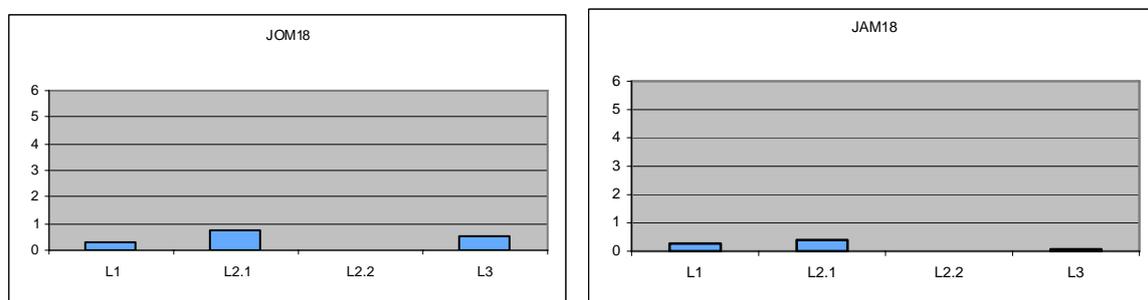


Figure 11. Number of AIS episodes per minute, divided by types for JOM (18 months) and JAM (18 months), both Thai children.

A difference between ALI and HAR reflected in Figure 10, but not matched by a corresponding difference between JOM and JAM in Figure 11, is a higher proportion of Level 2.2 episodes. We have thus far not discussed differences between Levels 2.1 (social) and 2.2 (interactive) co-ordinated AIS episodes in the data. Recall that the second, but not the first involves *attention-turning* on the part of the sender (infant), cf. sections 2 and 3. The theoretical motivation for this distinction is the following: In the case of social co-ordinated AIS, the infant tries to influence the adult by his own manifest behaviour, without checking whether the adult is attending. In contrast, in interactive co-ordinated AIS the infant attends to the adult's attentional state, a behaviour that, as argued in Brinck (in press b), might reflect an understanding of his having the adult's attention as a *prerequisite* for successful communication. If this is so (in general), then we may ask whether the model should not be

seen as involving *four*, rather than three distinct levels.⁸ Support for this would be data in favour for the *extra hypothesis* that the 18-month-old children would have a higher proportion of Level 2.2 to Level 2.1 compared to the 12-month group. However, the data rather showed the inverse pattern, as can be seen in Figure 12: The ratio of simple-to-interactive CAIS episodes was higher for the 18-month than for the 12-month old group.

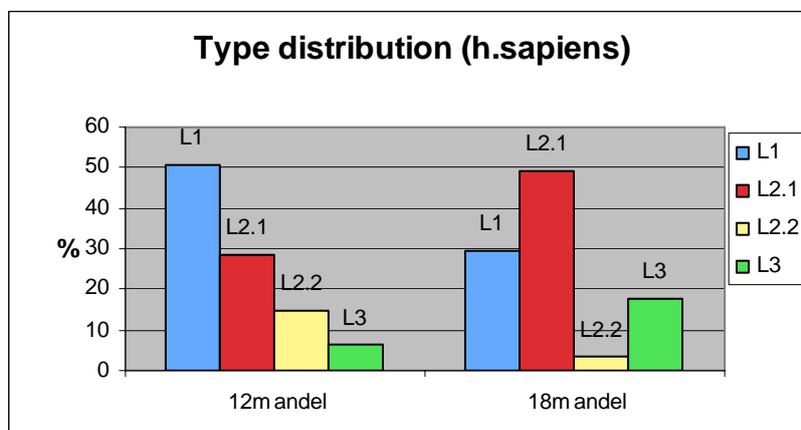


Figure 12. Percentages of different types (and sub-types) of AIS episodes for the 12-month and 18-month human children, showing a higher L1/L2 ratio for 18-month group.

At the same time, the unexpectedly high proportion of “simple” CAIS episodes in the 18-month group could reflect a problem with our definitions of Level 2.2. – in particular with respect to children who have begun to use language. The reason is that on many occasions these children pointed to a target and verbalized (sometimes even using the appropriate term in referring to an object, e.g. CHE pointing to a cartoon figure in a book and saying “Woodie!”). Since there was (in this and similar cases) no attention-turning towards the adult, the episode was coded as an instance of Level 2.1. However, it is likely that language by this age is taking the place of preverbal attention-getting behaviour. Moreover, if one takes the whole situational context into account – the parent, child and guest sitting on the floor, engaging in “naming games” with respect to toys and pictures – it is hardly surprising that the child does not check to see if the adult is paying attention. The interaction fills the pattern of many similar ones in the activity, and in a way, the child can *take it for granted* that the adult is paying attention – in the absence of evidence to the contrary. One could say that the attention of the other is part of the *common ground* (Clark 1996), and indeed, it seemed that the children directed attention to the adults whenever it, for some reason, e.g., a silence on the part of the adult, was not clear to the child that the adult was paying attention. Thus in sum, our particular coding scheme and data does not support treating Level 2.1 and Level 2.2 as developmentally distinct.

While it was part of our methodology to be ‘conservative’, as repeatedly pointed out, it seemed in quite a number of cases that we were forced to ‘underinterpret’ the children’s behaviour, since we had decided beforehand not to permit language to be coded as “referential behaviour” (cf. Section 3). Thus a number of /dæ/ utterances by TEA at 12 months which *appeared* to be referential (approximating the neuter demonstrative pronoun in Swedish), yet were not coded as such, and thus the corresponding AIS events were coded as Level 1 rather

⁸ A four-tiered model for intersubjectivity and co-operation is used in Brinck & Gärdenfors (2003)

2006-08-08

Level 2. It is possible that we have been too conservative in this case, since /dæ/ can be argued to be referential not only because of the corresponding demonstrative pronoun, but because it is formed by a protrusion of the tongue which is analogous to pointing, and has been argued by some to be even a developmental precursor to it (e.g. Williams 1992). It is characteristic that this data point, of 23 minutes, did not include a single case of true pointing (though a few referential reaching and waving movements were coded as referential behaviour). Clearly this is a topic that needs to be further investigated.

5. Summary and conclusions

In this report we have offered an *analysis of attentional intersubjectivity* (AIS) in terms of two types (symmetric and asymmetric) and three levels (synchronous, co-ordinated and reciprocal). We concentrated on the asymmetric type, and showed how the three levels formed a hierarchy of increasing complexity, with each successive level including additional behaviours on the part of the ‘sender’, and matching responses in the ‘receiver’. Interpreting the sender as the infant who initiates the AIS episode, we showed how these levels can be given a developmental interpretation, corresponding to a sequence of cognitive stages of development, based on the child’s understanding of the attentional state of the interacting adult. In sum, the foremost contribution of our analysis to ‘joint attention’ research concerns our systematic specification of the most important kinds of attentional intersubjectivity in terms of observable behaviours within one global model.

The model resulted in a coding scheme based in operational definitions of intersubjective behaviours, which was applied to infant-adult interactions in great apes (two chimpanzees and one bobobo), and human beings, divided in two age groups (12 and 18-month-old children). The analysis showed conspicuous differences between the two species (*Pan* and *Homo sapiens*), which we interpreted to be qualitative, and therefore as a possible contribution to the major goal of the SEDSU project: defining the “what it means to be human”. Still, since we concentrated on the visual modality (audio data from the apes was in practice unavailable), we need to take the results of our pilot study with some precaution. On the whole, however, the study offered support to our developmental interpretation of the different levels of AIS. The three hypotheses formulated prior to data analysis were supported, though it should be pointed out that the definitions of the levels, both the more general ones in Section 2 and the operational ones in Section 3, were further specified along the lines of Brinck (in press a) after a preliminary analysis of the data. At the same time, changes to the definitions were by no means introduced in order to offer *post hoc* support for our hypotheses, but in order to be able to code the interactions as unambiguously as possible. This forced us to be rather conservative in our definitions, excluding the children’s utterances as a form of referential behaviour, in order not to privilege the verbal children against the pre-verbal ones and the apes. The downside of this is that our model underestimates the role of language, as well as the capacity for sharing a ‘common ground’ without overt indications of this. The upside is that we managed to define the different levels in terms of observable behaviours only, which we view as an achievement in a field which is rife with debates on “rich” versus “lean” interpretations of the underlying capacities.

In conclusion, our model of attentional intersubjectivity, building on and further developing our previous research, can be said to have passed the test of empirical assessment, and will hopefully prove to be a useful conceptual and theoretical tool for further analyses. The most obvious further application of the model is to longitudinal developmental data such as that collected for Deliverable 15 within the SEDSU project. We hope to be able to do so for the theoretical summary of Work Package 6, *Intersubjectivity and Conventions*.

Acknowledgements

We wish to express our gratitude to Ulla Richthoff, Mathias Osvath, and Josep Call for allowing us to analyze the data which they had previously collected.

References

- Bard, K. (1992) Intentional Behaviour and Intentional Communication in Young Free-Ranging Orangutans. *Child Development*, 63, 1186-1197.
- Bard, K. (2005) Emotions in chimpanzee infants: the value of a comparative developmental approach to understand the evolutionary bases of emotion. In J. Nadel & D. Muir (Eds.), *Emotional Development*, pp. 31-60. Oxford: Oxford University Press.
- Brinck, I. (2001) Attention and the evolution of intentional communication. *Pragmatics and Cognition*, 9/2: 255-272.
- Brinck, I. (2003) The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly*, 3(4): 429-446.
- Brinck, I. (2004) Joint attention, triangulation and radical interpretation: A problem and its solution, *Dialectica*, 58(2): 179-206.
- Brinck, I. (in press a) The origin and essence of linguistic reference. *Dialectica*.
- Brinck, I. (in press b) Intersubjectivity and intentional communication. In J. Zlatev, T. Racine, C. Sinha and E. Itkonen (eds.) *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins.
- Brinck, I. and Gärdenfors, P. (2003) Co-operation and communication in apes and humans. *Mind and Language*, 18(5): 484-501.
- Butterworth, G. (2003) Pointing is the royal road to language for babies. In S. Kita (ed.) *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ: Laurence Erlbaum.
- Kaminski, J., Tiedel, J., Call, J., and Tomasello, M. (2005) Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object choice task. *Animal Behaviour*, 69, 11-18.
- Clark, H. (1996) *Using Language*. Cambridge: Cambridge University Press.
- Driver J., Davis G., Ricciardelli P., Kidd P., Maxwell E., and Baron-Cohen S. (1999) Gaze perception triggers reflexive visuospatial orienting. *Visual cognition*, 6(5), 509-54.
- Chawarska, K., Klin, A., and Volkmar, F. (2003) Automatic attention cueing through eye movement in 2-year-old children with autism. *Child Development*, 74(4), 1108-1122.
- Jackendoff, R. (1990) *Semantic Structures*. Cambridge, MA: MIT Press.
- Leavens, D. A., Hopkins, W. D. & Bard, K. (1996) Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 110: 346-353.
- 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: 417-425.
- Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005) Intentionality as Measured in the Persistence and Elaboration of Communication by Chimpanzees (*Pan troglodytes*). *Child Development*, 76, 291-30
- Moore, C. and Dunham, P. (1995) *Joint Attention: Its Origins and Role in Development*. Hillsdale, NJ.: Lawrence Erlbaum.
- Tomanoga, M., Myowa-Yamakoshi, M., Mizuno, Y., et al. (2004). Development of social cognition in infant chimpanzees (*Pan troglodytes*): Face recognition, smiling, gaze, and the lack of triadic interactions. *Japanese Psychological Research*, 46(3), 227-235.
- Tomasello, M. (1999) *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28: 675-735.

- Williams, S. (1992) *A Study of the Occurrence and Functions of "da" in a Very Young Bilingual Child*. Verlag an der Lottbek.
- Zlatev, J. (2003) Meaning = Life (+ Culture). An outline of a unified biocultural theory of meaning. *Evolution of Communication*, 4/2: 253-296.
- Zlatev, J. (in press) Intersubjectivity, language, and the mimesis hierarchy. In J. Zlatev, T. Racine, C. Sinha and E. Itkonen (eds.) *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins.
- Zlatev, J., Persson, T. & Gärdenfors, P. (2005). *Bodily mimesis as "the missing link" in human cognitive evolution*. Lund: Lund University Cognitive Studies.