

8 Stages in the Development of Perceptual Intersubjectivity

Jordan ZLATEV, Ingar BRINCK, Mats ANDRÉN

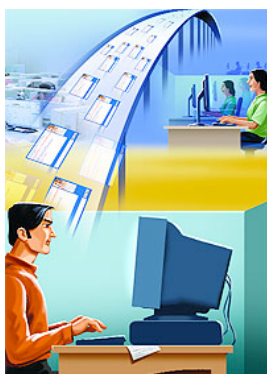
Abstract. We offer a model of perceptual intersubjectivity (PI), the phenomenon of *two or more subjects focusing their attention on the same external target*. The model involves two types: symmetric and asymmetric PI, and three levels: synchronous (SPI), coordinated (CPI) and reciprocal (RPI), defined on the basis of the observable behavior of the participants of (non-verbal) social interactions. We hypothesize that the three levels correspond to *stages* in the development and possibly evolution of human perceptual intersubjectivity, and provide support for this through an empirical study of adult-infant interactions in two species of great apes (chimpanzees and bonobos) and human beings. The results showed conspicuous and apparently qualitative differences between the human and non-human subjects, and clear developmental patterns in the human data. Thus our analysis may contribute to the ultimate goal of understanding the nature and development of human cognitive specificity, in line with goals with the collaborative project *Stages in the Evolution and Development of Sign Use* (SEDSU) [1].

Contents

8.1	Introduction.....	118
8.2	A model of perceptual intersubjectivity.....	118
8.3	An empirical study.....	125
8.4	Conclusions.....	131
8.5	Acknowledgment.....	132
8.6	References.....	132

Support us – Ask your institution/library/department to order our books or to apply for series membership - ISSN 1566-7677

To order a book
contact by fax or
by e-mail the offices
below:



IOS Press, Inc.
4502 Rachael Manor
drive
Fairfax, VA 22032
U.S.A.
Tel.: +1 703 323 5600
Fax: +1 703 323 3668
sales@iospress.com

IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
The Netherlands
Tel: +31 20 688 3355
Fax: +31 20 620 3419
info@iospress.nl

IOS Press,
c/o Ohmsha, Ltd.
3-1 Kanda Nishiki-cho
Chiyoda-ku
Tokyo 101
Japan
Fax: +81 3 3233 2426
(Books only)

IOS Press/
Gazelle Book Services Ltd
White Cross Mills
Hightown
Lancaster LA1 4XS
United Kingdom
Tel.: +44 1524 68765
Fax: +44 1524 63232
sales@gazellebooks.co.uk

Enacting Intersubjectivity

A Cognitive and Social Perspective on the Study of Interactions

Volume 10 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: F. Morganti, A. Carassa and G. Riva

May 2008, approx. 280 pp., hardcover

ISBN: 978-1-58603-850-2 **NEW**

Price: US\$161 / €115 / £81

From Communication to Presence

Cognition, Emotions and Culture towards the Ultimate Communicative Experience

Festschrift in honor of Luigi Anolli

Volume 9 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: G. Riva, M.T. Anguera, B.K. Wiederhold, F. Mantovani

September 2006, 323 pp., hardcover

ISBN: 978-1-58603-662-1

Price: US\$161 / €115 / £81

Global Data Management

Volume 8 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: R. Baldoni, G. Cortese, F. Davide and A. Melpignano

July 2006, 376 pp., hardcover

ISBN: 1-58603-629-7

Price: US\$161 / €115 / £81

The Hidden Structure of Interaction

From Neurons to Culture Patterns

Volume 7 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: L. Anolli, G. Riva, S. Duncan Jr. and M.S. Magnusson

May 2005, 304 pp., hardcover

ISBN: 1-58603-509-6

Price: US\$161 / €115 / £81

Ambient Intelligence

The Evolution of Technology, Communication and Cognition Towards the Future of Human-Computer Interaction

Volume 6 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: G. Riva, F. Vatalaro, F. Davide and M. Alcañiz

January 2005, 316 pp., hardcover

ISBN: 1-58603-490-1

Price: US\$161 / €115 / £81

Being There

Concepts, Effects and Measurements of User Presence in Synthetic Environments

Volume 5 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: G. Riva, F. Davide and W.A. IJsselstein

2003, 344 pp., hardcover

ISBN: 1-58603-301-8

Price: US\$161 / €115 / £81

Say not to Say: New Perspectives on Miscommunication

Volume 3 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: L. Anolli, R. Ciceri and G. Riva

2001, 288 pp., hardcover

ISBN: 1-58603-215-1

Price: US\$161 / €115 / £81

Towards CyberPsychology

Mind, Cognition and Society in the Internet Age

Volume 2 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: G. Riva and C. Galimberti

2001, 326 pp., hardcover

ISBN: 1-58603-197-x

Price: US\$161 / €115 / £81

Communications Through Virtual Technologies

Identity, Community and Technology in the Communication Age

Volume 1 Emerging Communication: Studies on New Technologies and Practices in Communication

Edited by: G. Riva and F. Davide

2001, 292 pp., hardcover

ISBN: 1-58603-162-7

Price: US\$161 / €115 / £81

8.1 Introduction

Intersubjectivity can be defined, most generally, as the *sharing of states and processes of consciousness between two or more subjects* [2]. Different forms of intersubjectivity can be distinguished on the basis of the most prominent form of consciousness involved – perceptual, affective, or reflective [3] – or on the basis of their intrinsic complexity [4-6]. An important value of such analyses is that they can help us understand the development and evolution of human intersubjectivity. The goal of the present chapter is to provide one such analysis, focusing on the phenomenon of *perceptual intersubjectivity*, more commonly known under the label ‘joint attention’. Though wide-spread, the meaning of the latter term is rather ambiguous: sometimes it refers to the general case in which two or more subjects perceive the same target [7], while for others the term applies to more specific reciprocal states in which the subjects are also *aware* that they perceive the same target [8]. 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 [5,6,9-15] we define perceptual intersubjectivity (PI) as the process in which *two or more subjects focus their attention on the same external target*. Like most others we focus on the visual modality, but formulate our definitions in way to be applicable to other modalities such as hearing and touch as well. In Section 2, we identify different levels of PI on the basis of the complexity of the interaction between the subjects. These levels build on each other cumulatively, and it is therefore possible to hypothesize that they correspond to developmental and/or evolutionary *stages* [1].

In order to test this hypothesis we provide *operational definitions* of the different levels of PI, thereby making the model applicable to empirical data involving human beings and non-human primates, and thus allowing it to be empirically assessed. One such an assessment is described in Section 3, where we apply the model to data from adult-infant interactions in great apes (chimpanzees and bonobos) and human beings. In Section 4 we summarize our proposed stage-based model of perceptual intersubjectivity.

8.2 A model of perceptual intersubjectivity

8.2.1 General Definitions

In most general terms perceptual intersubjectivity (PI) can be defined as the phenomenon of *two or more subjects focusing their attention on the same external target*. Individual PI episodes may be individuated in terms of their targets, present in the immediate context shared by the participants of the interaction. 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). The term ‘object’ should be understood in a wide sense to refer to any animate or inanimate entity that occupies a position in space-time, e.g. a toy, or a person. A PI episode may have one and only one target¹. Two major types of PI episodes can be distinguished:

¹ See Section 3.2 for a more precise operational definition of the notion ‘PI episode’

- **Symmetric:** when the target has already been noticed by both (or more) subjects [6]. In discussing symmetric PI, we refer to the participants as subjects.
- **Asymmetric:** when the target initially is noticed by *only one* of the subjects, and subsequently the other subject aligns her attention with the first subject's attention [9- 12]. We refer to the subject who has initially noticed the target as the sender and the one who focuses her attention on the target as a consequence of the sender's behavior the receiver.²

The relationship between these two types merits further research, but this is not our present focus. Rather, we will concentrate on the asymmetric type, especially in the study described in Section 3. Nevertheless, we maintain that the two types have *parallel* levels or stages, and we capture this parallelism in the presentation below.

Prior to describing the different levels of symmetric and asymmetric PI, we provide definitions of the central terms that will appear in the descriptions. The following four are basic behaviors, which can be observed more or less directly:

- **Attention-focusing:** the sender's or receiver's prolonged attention to a target;
- **Attention-turning:** the sender's change in attention-focusing from target to receiver (or vice versa); in the case of visual attention this amounts to gaze alternation;
- **Attention-getting:** the sender's behavior directed at the receiver, apparently causing the receiver to turn her attention toward the sender;
- **Attention-contact:** the sender's and receiver's focused attention on each other's attentional state; in the case of visual attention this amounts to mutual gaze.

On this basis, we can define the following more complex behaviors:

- **Referential behavior:** the sender's behavior while attention-focusing on the target, apparently causing the receiver to turn her attention toward the target. Referential behavior can be either:
- **Communicative:** performed relative to the attentional status of the receiver, with the goal of affecting her behavior; or
- **Non-communicative:** performed in order to manipulate an object, reach a location etc. and not with the goal of affecting the receiver's behavior.

Non-communicative referential behavior may of course be "communicative" from the perspective of the receiver, but it is not intentionally communicative for the sender. Intentional communication [16] is indispensable for higher levels of intersubjectivity, and cannot be reduced to observable behavior or described in purely causal terms –short of behaviorism, which as well-known is a blind alley. On the other hand, intentional communication is not something "private" and unobservable, since the intentions it involves are behaviorally manifest. Attention-getting, attention-turning and persistence are markers of intentional communication, as stated in a classical definition:

² 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 refer to the sender in the masculine and the receiver in the feminine.

“Intentional communication is a signalling behavior in which the sender is aware, a priori, of the effect that the signal will have on his listener, and he persists in that behavior until the effect is obtained or failure is clearly indicated. The behavioral evidence that permits us to infer the presence of communicative intentions includes (1) alternation in eye gaze contact between the goal and the intended listeners, (2) augmentations, additions, and substitution of signals until the goal has been obtained, and (3) changes in the form of the signal towards abbreviated and/or exaggerated patterns that are appropriate only for achieving a communicative goal.” [17 p. 39]

Other more specific behavioral manifestations apply to the most characteristic form of non-verbal intentional communication, pointing, which can be distinguished from its non-communicative (in the above sense) counterpart reaching as follows:

- **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 – in order to affect another subject’s behavior towards the target.
- **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. The action is performed irrespective of the attention of another subject.

Although pointing gestures may look similar to reaching, the two can be distinguished on closer examination. If the sender is not persistently trying to decrease the distance between himself and the target, and the reach and the grip of the hand are not adjusted so as to fit the target, then this should be classified as an instance of pointing (by the definition above), rather than reaching.

8.2.2 Levels of perceptual intersubjectivity

Based on the definitions given above, we distinguish three general levels of PI, 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 a communicative behavior, in the sense that the action is not performed with the goal of affecting the behavior 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 [5]. In the remaining part of this section we specify more clearly each one of these three levels, with respect to the two major types of PI: symmetric and asymmetric.

Level 1: Synchronous Perceptual Intersubjectivity (SPI)

In the case of symmetric SPI, the target T has independently captured subject A’s and subject B’s attention and caused both to focus their attention on it (Figure 1). An example would be when T belongs to a category of similar intrinsic value for both A and B, such as food or danger [4].

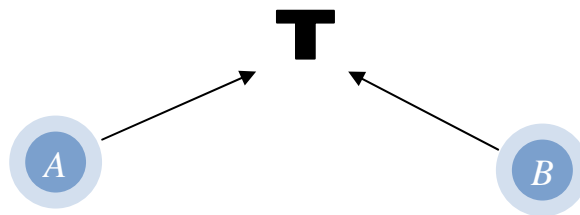


Figure 1. Symmetric SPI

On the other hand, asymmetric SPI can be characterized by the following stereotypical sequence (Figure 2):

1. A focuses his attention on T, possibly reaching towards T.
2. B’s attention is attracted by (1).
3. B turns her orientation to T, with the result that both A and B focus their attention on T.

An example is an infant’s reaching toward T, causing a caregiver to notice T, and possibly to offer it to an infant. But even more simple behaviors without reaching, such as *attentional contagion*, would qualify as belonging to this level, e.g. when a goat A turns its attention to a significant target (food) located behind another goat B, and this causes B to look towards A, and then turn its attention to the target by following the direction of A’s attention [18].

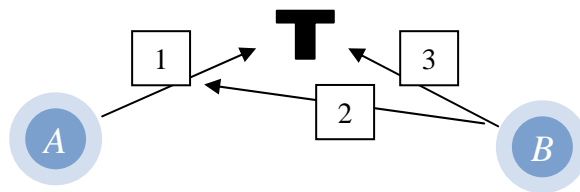


Figure 2. Asymmetric SPI

Level 2: Coordinated Perceptual Intersubjectivity (CPI)

In the case of symmetric CPI, T has already been noticed by A and B. In addition, A directs his attention to B’s attention-focusing on T, and B directs 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 (Figure 3).

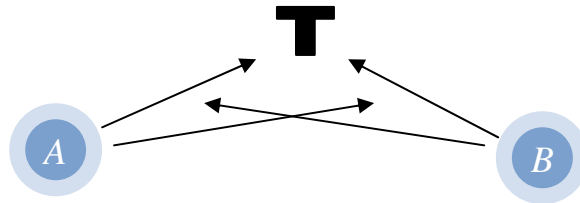


Figure 1. *Symmetric* coordinated perceptual intersubjectivity (CPI)

An example of symmetric CPI is the following situation of *social referencing*: A is an infant, B is an adult, and 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 attitude to T. Further, by monitoring A's attention and attitude towards T, B can check if A is behaving appropriately towards T.

Asymmetric CPI can be divided into two sub-types, according to whether or not the sender turns his attention to and focuses on the receiver during the interaction: (a) simple CPI in which the sender ostensibly attends to the target and engages in communicative referential behavior towards it, however, without turning his attention to the receiver; (b) complex CPI where the sender turns his attention to the receiver, and possibly draws her attention to himself and his behavior (attention-getting). Thus, simple CPI can be characterized by the following stereotypical sequence of behaviors (Figure 4):

1. A focuses his attention on T.
2. A ostensibly attends to T and engages in communicative referential behavior towards T.
3. B notices (2).
4. B turns her attention to T.

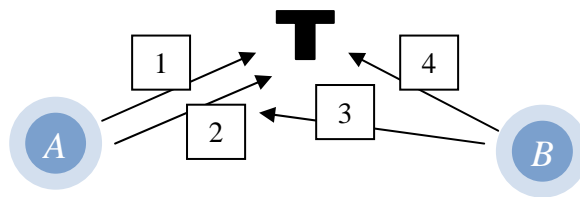


Figure 4. *Asymmetric simple* coordinated perceptual intersubjectivity (CPI)

Instances of simple CPI are *simpler forms of imperative pointing*, where an infant points toward an object, without turning his attention to the adult. Note that the novel behavior that distinguishes simple coordinated from synchronous PI is step 2, the communicative referential behavior, e.g., manifest attention (e.g., gaze) or pointing.

Complex CPI may be characterized by the following stereotypical sequence (Figure 5):

1. A focuses his attention on T.

- 2a. A ostensibly attends to T and engages in communicative referential behavior towards T.
- 2b. A turns his attention towards B.
3. B notices (2).
4. B turns her attention to T.

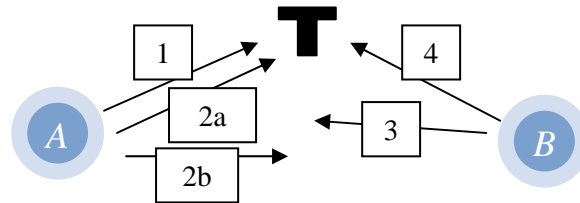


Figure 5. *Asymmetric complex* coordinated perceptual intersubjectivity (CPI)

Examples of this are *typical cases of imperative pointing*, in which a child “makes sure” that the adult is attending before performing the pointing gesture. Note that the crucial behavior that distinguishes complex from simple CPI is step 3 in which the infant turns his attention towards the adult during the interaction.

Level 3: Reciprocal Perceptual Intersubjectivity (RPI)

In symmetric RPI, A not only attends to B’s attention to T and vice versa (as in symmetric CPI), but attends to B’s attending to his (A’s) attention, and vice versa. On this level we have *third-order attention* [6]. The following example may illustrate the phenomenon: A child and an adult play a game of 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-level attention of the child (A) is shown.

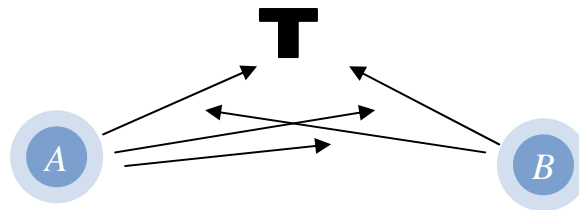


Figure 6. *Symmetric* reciprocal perceptual intersubjectivity (RPI)

Asymmetric RPI is characterized by the following stereotypical sequence (Figure 7):

1. A focuses his attention on T.
2. A engages in attention-getting relative to B (optional).
3. B notices (2), and focuses her attention on A.
4. A and B establish attention contact.
5. A ostensibly attends to T and/or engages in communicative referential behavior towards T.
6. B turns her attention to T.

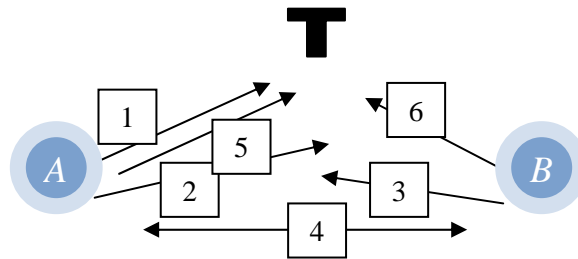


Figure 7. Asymmetric reciprocal perceptual intersubjectivity (RPI)

Examples of asymmetric RPI are typical cases of *declarative pointing*. Note that the novel behavior, distinguishing RPI from CPI is step 4, attention-contact, which in the case of visual attention corresponds to mutual gaze.

8.2.3 Summary

In this section we have provided a level-based analysis of perceptual intersubjectivity, where each consecutive level is of higher *complexity* than the previous one. In the case of symmetric PI, 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 PI 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 referential behavior of the first, the different levels are defined by sequences of behaviors. In this type of PI the increased complexity is reflected by the fact that each higher level *subsumes* the previous ones, and also includes crucial novel behaviors.

Thus, on Level 1 the behavior of the sender is not (intentionally) communicative (it is not directed towards the attention of the receiver). In contrast, on Level 2 (coordinated PI) the sender engages in various forms of communicative referential actions such as pointing. On Level 2.1 the sender engages in ostensibly manifest behaviors, but does not turn his attention to the receiver to check if his action has been noticed. In contrast, on Level 2.2 such attention-turning occurs. Finally, Level 3 adds attention-contact, during which the subjects simultaneously attend to each other's attentional states, which in the visual modality corresponds to mutual gaze.

The different types and levels of perceptual intersubjectivity may thus be distinguished on the basis of overt behaviors and their sequencing. The rationale behind this procedure is its purpose, i.e., to construct a global model of perceptual intersubjectivity that can be applied to empirical data of adult-infant interactions in different species, cultures, and at different ages, such as the data made available by the SEDSU project [1]. This is mandatory in order to substantiate our hypothesis that the levels that we have identified 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 the study described in the following section.

8.3 An empirical study

8.3.1 Data and hypotheses

We analyzed the following sets of data, which were made available by our collaborative 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 a Thai/Swedish video-linked corpus, involving three Swedish children (BEL, TEA and HAR) and 3 Thai children (JOM, JAM and CHE), when these children were app. 18 months old.
- 2 video-recording of 2 Swedish children aged 12 months: ALI (recorded by Mats Andréén, 15 minutes) and TEA (recorded by Ulla Richtoff, 23 minutes).

As obvious from this description, with the exception of the 2 data points from TEA and the bonobo, the data was not longitudinal. However, given that the data was, in broad terms, cross-sectional, and the different PI levels are of increasing complexity, we could formulate three hypotheses to test whether these correspond to developmental and evolutionary stages. H1: PI episodes of Level 3 (Reciprocal PI) will be attested predominantly among the 18-month old children. H2: PI episodes of Level 2 (Coordinated PI) will be observed among the 18-month old children and the 12-month old children. H3: PI episodes of Level 1 (Synchronous PI) will be the only form of perceptual intersubjectivity found in the ape data – perhaps with occasional instances of Level 2.

8.3.2 Operationalization of the model

The definitions of the different types and levels of PI presented in Section 2 were intended 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* for the study it was necessary to specify them further. The general guiding principle was to be *conservative*, i.e. to have operational definitions which preferably under-interpret rather than over-interpret the observational data, especially the data from the 18-month old children. The reason for this is that these children have already made their entrance into language, and language can *substitute* for many other forms of intersubjective behavior, including mutual gaze and gesturing [7-8]. While being in essence a form of communicative referential behavior we decided explicitly not to code the children's utterances as such, since that would have placed them on an uneven footing compared to the behavior of pre-linguistic children and apes. Furthermore, especially with the 12-months old children, 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 behaviors (*attention-turning, attention-contact, communicative referential behavior*).

The first specification compared to the definitions of asymmetric PI in Section 2, is that we analyzed only cases in which the sender was the *infant* (human or ape) and the receiver was 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 PI, the main reason being that it was much easier to individuate the PI episodes for this type, rather than for symmetric type.

The beginning of a new PI episode was marked by the introduction of a new target: an object, event, location or direction that received focal attention from one or both of the subjects. Therefore, a PI episode by definition includes one and only one target and the introduction of another target defines the end of the previous episode and the start of a new one. What operationally counts as a new target was based on *visible behavioral contrasts* in the interaction. New targets were judged to occur when:

- (1) There was a shift in the infant's attention to a target that is altogether outside his earlier focus of attention.
- (2) There was a shift in the infant's attention to a target which is more or less within the earlier focus of attention, but a visible shift of attention is observable in both the infant's and the adult's behavior, such as:
 - a. *An object is singled out in contrast to several possible others.*
Example: playing with building blocks; although both participants are already attending to the block-building in general, the infant focuses his attention on a specific block while picking it up and thereby introduces 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 in contrast to the object as a whole.*
Example: playing with a toy telephone; although the toy telephone is already the focus of the infant's attention, the infant shifts his attention to a specific part 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 moved and in this process its new location constitutes a new target of attention.* Example: the infant is holding a glass of milk and the glass is within his focus of attention; but then the infant puts it down outside the current visual field of the adult who then needs to shift attention to this new location.

Since we were primarily interested in classifying the behavior of the infant, when new targets were altogether outside the focus of the infant's previous attention (case 1) it was not of crucial importance whether the target was within the visual field of the adult as long as the infant cannot see this. However, in cases (2a), (2b) and (2c) it was of crucial importance whether there was also a slight adjustment in the attention of the adult. Otherwise it was impossible to establish that these more subtle kind of new targets really are established as common to both parties.

Since it was not possible to distinguish the infant's attention to the target from attention-turning when the target was the adult herself, these targets/episodes were

excluded from analysis. In other words, the analysis deals with triadic and not dyadic engagements. Finally, we concentrated on visual attention, thus reducing attention to “gaze”. In the case of Level 3 (RPI) we requested that the infant and adult engage in mutual attention (gaze) – *within the infant’s turn*, i.e. prior to the adult verbally commenting on the target. Given these qualifications, we could operationally define the different levels of asymmetric PI as shown in Table 1.

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

Table 1. Operational definitions of the levels of perceptual intersubjectivity used for the empirical study

8.3.3. Analysis and results

All PI episodes were identified in the data according to the criteria outlined above. Coding was performed by the first and third author, and in case of uncertainty, the second author was consulted as well: until consensus between the three authors was reached. Unclear examples were excluded. This resulted in a total of 190 PI episodes, divided by the different video-recordings (‘data points’) as shown in Table 2. The results strongly suggest that asymmetric PI seems to be a human speciality – not in the sense that it is *unique* for our species, but that it is much more frequent in human 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.

Data point	Age (months)	Length (minutes)	# PI episodes
Kara (chimp)	7	8:30	2
Lobo (chimp)	19	60	1
Luiza (bonobo)	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

Table 2. Total number of PI episodes per data point

Furthermore, the three hypotheses (Section 3.1) were almost surprisingly well confirmed, as shown in Figure 8. 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 were 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), thereby confirming hypothesis 3.

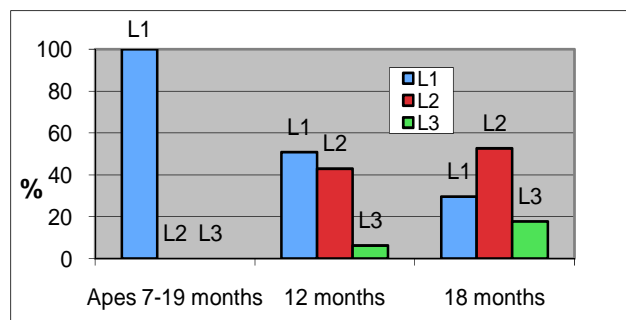


Figure 8. Percentage of PI episodes by level and age group. Total number of episodes: 5 for apes, 50 for 12-month old children, and 135 for 18-month old children

8.3.4. Discussion

The results of the study supported our hypothesis that the different levels of PI in our model, at least of the asymmetric variety, 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 imply more complex processes than *attentional contagion*.³ Of course, this does not exclude the possibility of higher-level processing.

The complete absence of any complex type 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 we know from previous research that in captivity (adult) apes do engage in intentional communication with human subjects [19]. Nevertheless, the differences were so conspicuous, that we believe that they reflect a qualitative difference in the nature of ape and human social interactions: human infants (and young children) engage in communicative referential behavior 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 8 for the two groups) this seems to be a feature of human social interactions that is *independent of and more basic than language*. In terms of the Mimesis Hierarchy (MH) model [14-15] this difference can be interpreted as due to differences in dyadic and – especially – triadic *bodily mimesis*.

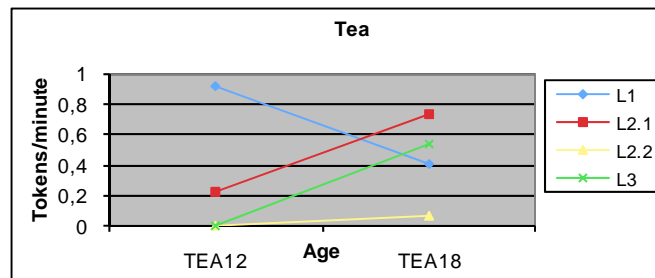


Figure 9. Number of PI episodes per minute for TEA at 12 and 18 months.

This conclusion is also supported by the only piece of direct human developmental evidence in our data: While SPI (Level 1) episodes predominate over all the others for the child TEA at 12 months, occurring roughly once a minute in the interaction, they decrease to 0,4 per minute at 18 months, while CPI (Level 2) episodes rapidly increase (Figure 9).

The results concerning our first hypothesis (H1) regarding RPI episodes dominating in the 18-month old group, the results were less clear cut. Indeed, there was a higher proportion of RPI episodes in that group, but the 12-months old

³ Attentional contagion appears to be supported by a specialised neural mechanism [20]. The attention system immediately reacts to the perceivable re-orientation of the body, head, or gaze, or all of these, of other subjects, and will cause the receiver of the signal to turn her attention unless the behaviour is inhibited. Thus, attentional contagion can be said to occur on a “subpersonal” processing level, i.e., a level that cannot be accessed by conscious awareness.

group had only 2 children, and there was considerable individual variation between the children in both groups. Also the distinction between “simple” and “complex” CPI on the basis of the presence of attention-turning (towards the adult) in the latter, but not the former, seemed questionable. As Figure 10 shows, there was a higher proportion of simple CPI (Level 2.1) in the older group.

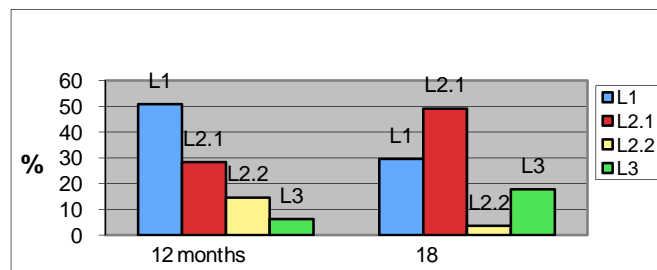


Figure 10. Percentages of different levels (and sub-levels) of PI episodes for the 12-month and 18-month human children

On many occasions the older 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 no attention-turning to the adult, the episode was coded as an instance of Level 2.1. But if one takes the whole situational context into account – the parent, child and guest sitting on the floor and repeatedly naming different toys and pictures – it is hardly surprising that the child does not check to see if the adult is paying attention, since the interaction conforms to a pattern of many similar ones, and in a way, the child can *take it for granted* that the adult is paying attention – given the lack of any evidence to the contrary. One could say that the attention of the other is part of the *common ground* [21]. Indeed, it seemed that the children directed their attention to the adults when for some reason, e.g. a silence on the part of the adult, it 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 pointed out, it seemed in quite a few cases that we were forced to under-interpret the children’s behavior, since language was not allowed to be coded as “communicative referential behavior”. Thus a number of /dæ/ utterances by TEA at 12 months which did appear to be communicative (and approximating the neuter deictic pronoun in Swedish), were not treated as such, and thus the corresponding PI episodes were coded as Level 1 rather Level 2. We may have been too conservative in this case, since /dæ/ can be argued to be communicatively referential not merely because of the corresponding deictic 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 [22]. It is characteristic that this data point (of 23 minutes) did not include a single case of true pointing (though a few cases of reaching). Clearly, this is a topic that needs to be further investigated.

8.4 Conclusions

In this chapter we have offered a model of perceptual intersubjectivity (PI) in terms of two types (symmetric and asymmetric) and three levels (synchronous, coordinated and reciprocal). We concentrated on the asymmetric type and showed how the three levels formed a complexity hierarchy, with each successive level including additional behaviors on the part of the sender and matching responses from the receiver. Interpreting the sender as the infant initiating the PI episode, we showed how these levels can be given a developmental interpretation, corresponding to a possible sequence of stages of development, distinguished by the child's progressive understanding of the attentional state of the interacting adult. In sum, the foremost contribution of this chapter to the theme of "enacting intersubjectivity" consists of the systematic specification of levels of perceptual intersubjectivity in terms of *observable behaviors* for the purpose of analyzing social interaction, thereby connecting the "individual" and the "social" dimensions.

The model resulted in a coding scheme of operational definitions, 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 results showed conspicuous differences between the two species (*Pan* and *Homo sapiens*), which we took to be qualitative, and therefore as a possible contribution to the ultimate goal of understanding human cognitive specificity. 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 study with some precaution.

Nevertheless, the results offered support to our developmental (and to some extent evolutionary) interpretation of the different levels of PI. The three hypotheses formulated prior to any data analysis were confirmed, though it should be pointed out that the definitions of the levels, especially the operational ones in Section 3, were further specified after 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 to be able to code the interactions as unambiguously as possible. This forced us to exclude the children's utterances as a form of communicative referential behavior, in order not to privilege the verbal children against the pre-verbal ones and the apes. The downside of this is that our model somewhat 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 behaviors. We view this as an achievement in a field which is rife with debate on "rich" versus "lean" interpretations of the underlying capacities.

In conclusion, our model of perceptual intersubjectivity, building on and further developing our previous research, can be said to have passed the test of empirical assessment, and can therefore be regarded as a useful conceptual and theoretical tool for conducting further analyses.

8.5 Acknowledgements

We wish to express our gratitude to Ulla Richtoff, Mathias Osvath and Josep Call for supplying us with parts of the data, and to the collaborative project *Stages in the Evolution and Development of Sign Use* (SEDSU), supported by the EU FP6 program under the call “What it means to be human”, for providing the framework and resources necessary for conducting this research.

8.6 References

- [1] J. Zlatev & SEDSU-Project, Stages in the Evolution and Development of Sign Use (SEDSU). In A. Cangelosi, A. D. M. Smith & K. Smith (Eds.), *The Evolution of Language: Proceedings of the 6th International Conference (EVOLANG6)*, (pp.379-388). New Jersey: World Scientific, 2006.
- [2] J. Zlatev, T. Racine, C. Sinha & E. Itkonen, *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam/Philadelphia : Benjamins, in press.
- [3] T. Honderich, Radical externalism. *Journal of Consciousness Studies*, 13 (7-8), 3-13, 2006.
- [4] S. Bråten, *On Being Moved: From Mirror Neurons to Empathy*. Amsterdam/Philadelphia : Benjamins, 2007.
- [5] I. Brinck, The role of intersubjectivity for the development of intentional communication. In Jordan Zlatev, et al. *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam : Benjamins, in press.
- [6] J. Zlatev, The co-evolution of intersubjectivity and bodily mimesis. In J. Zlatev, et al. (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam : Benjamins, in press.
- [7] G. Butterworth, Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ : Laurence Erlbaum, 2003.
- [8] M. Tomasello, *The Cultural Origins of Human Cognition*. Cambridge, Mass. : Harvard University Press, 1999.
- [9] I. Brinck, Attention and the evolution of intentional communication. *Pragmatics & Cognition*, 9 (2), 255-272, 2001.
- [10] I. Brinck, The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly*, 3/4, 429-446, 2004.
- [11] I. Brinck, Joint attention, triangulation and radical interpretation: A problem and its solution. *Dialectica*, 58 (2), 179-205, 2004.
- [12] I. Brinck & P. Gärdenfors, Co-operation and communication in apes and humans. *Mind & Language*, 18 (5), 484-501, 2003.
- [13] J. Zlatev, Meaning = Life + (Culture): An outline of a biocultural theory of meaning. *Evolution of Communication*, 4 (2), 253-296, 2003.
- [14] J. Zlatev, T. Persson & P. Gärdenfors, *Bodily Mimesis as the "Missing Link" in Human Cognitive Evolution, LUCS 121*. Lund : Lund University Cognitive Studies, 2005.
- [15] J. Zlatev, From protomimesis to language: evidence from primatology and social neuroscience. *Journal of Physiology*, Paris, in press.
- [16] P. Grice, Meaning. *Studies on the Way of Words*, (pp. 213-223). Harvard : Harvard University Press, 1989.
- [17] E. Bates, *The Emergence of Symbols. Cognition and Communication in Infancy*. New York : Academic Press, 1979.
- [18] J. Kaminski, J. Tiedel, J. Call & M. Tomasello, Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object choice task. *Animal Behaviour*, 69, 11-18, 2005.
- [19] D. A. Leavens & W. D. Hopkins, The whole hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology*, 113 (4), 417-425, 1999.
- [20] K. Chawarska, A. Klin & F. Volkmar, Automatic attention cuing through eye movement in 2-year old children with autism. *Child Development*, 74 (4), 1108-1122, 2003.
- [21] H. Clark, *Using Language*. Cambridge : Cambridge University Press, 1996.
- [22] S.A. Williams, *Study of the Occurrence and Functions of "da" in a Very Young Bilingual Child*. Lottbek : Verlag an der Lottbek, 1992.