FINO

Northwestern Italian Philosophy Consortium

University of Genoa – University of Turin

Learning Through Others.

Natural pedagogy and Mindreading:

A possible cooperation

PhD Thesis

XXX Ciclo

Candidate:

Emiliano Loria

Matricola: 808122

Tutor:

Prof. Cristina Meini

Evaluators:

Prof. Pietro Perconti

Prof. Carlo Lai

Turin, 2018

To my son, source of joy and knowledge.

He moves me something special, every day.

Learning Through Others.

Natural pedagogy and Mindreading: a possible cooperation

Introduction.

1. Eating a soup 7

1.1. The man who wanted to teach how to eat a soup with the spoon 8

2. Natural pedagogy theory in a few words 9

2.1. The supposed independence of natural pedagogy from mindreading 11

3. Aims of the present research 12

Chapter 1. ***Action interpretation criteria for infants' mind***

1. Communicative and instrumental acts 15

2. Reading actions 16

2.1 "Action-oriented paradigm" in infant cognition 16

3. Types of action and types of agent: a useful distinction 19

4. Instrumental action (interpretation) 20

4.1 Teleological stance and mindreading overlapping 22

4.1.1 The independence of the two interpretational schema in Psychopathology 24

4.2 Naïve theory of rational action against motor mirroring mechanism for action understanding 25

4.2.1 Southgate and colleagues' experiments in favour of the application of the principle efficiency in instrumental action interpretation 27

4.3 Neuronal correlates for teleological reasoning 29

4.4 Neural correlates for referential expectation in communicative actions: the Pomiechowska and Csibra's proposal 29

5. Context and neural mechanism for action interpretation 32

6. Communicative context: the heritage of the Relevance theory for the natural pedagogy 33

6.1 Recognition of communicative intention: some differences and similarities between Gricean account and communicative principle of relevance 35

6.2 *Neo-Gricean recipe* 40

7. Communicative action (interpretation) 42
7.1 Second person view: some reflections 43

Chapter 2. ***The infant sensitivity to ostensive communication***

1. Communicative faces 47
1.1 Newborns prefer communicative faces 48
1.2 What mechanism is involved for early gaze detection? 51
1.2.1 Adults and infants: similarity or not similarity? 53
1.3 Tactile communication in newborns 55
2 Referential expectations
2.1 Eye contact influences the processing of objects in 4-month-old infants 58
2.2 Eye contact influences the processing of objects in 5- month-old infants 59
2.3 "There is something for me!" 60
2.3.1 Responses to multimodal ostensive signals 61
2.3.2. Marno and colleagues' study on referential expectation in 4-month-old 66
3 TOM and the trigger of natural pedagogy system 67
4 An innate epistemic principle? 69

Chapter 3. ***The Natural Pedagogy theory: «A relevance-guided learning device»***

1. Precursor studies of natural pedagogy theory 73
1.1 Rational imitation explanation 74
1.2 Replications of Gergely and colleagues' experiment 76
1.2.1 Sensitivity to ostensive cues: comparing behaviours between distantly related species 77
1.2.2 Paulus and Bekkering's studies 78
1.2.3 An integrative model 80
1.3Meltzoff's Like-me hypothesis 81
1.4 Conceptualising MNs: hypothesises 84
1.5 MNs and infant imitation 84
1.6 Tomasello et al.'s theory of infant social learning 86
2. Natural pedagogy theory 87
2.1. Functions of ostensive cues in natural pedagogy 89
2.2 The Epistemic trust triggered by ostensive cues 90
2.3 Assumption of generalizability 92

2.3.1 Futó, Téglás, Csibra, and Gergely's experiment 94
2.3.2 Generalizability in preschoolers: Butler and Markman's experiment 97
2.4 Assumption of universality 101
2.4.1 Gergely and colleagues' experiment (2007) 103
2.4.2 Egyed and colleagues' experiment about preferences socially shared 106
2.5 Universality without mindreading? 108
3. Investigating natural pedagogy in pre-schooler children 110
3.1 Sage and Baldwin's experiment 110
4. Natural pedagogy in a teaching evolutionary perspective 112
5. Evolutionary origins of natural pedagogy 116
6. Natural pedagogy is universal 118
7. Natural pedagogy and psychological essentialism 118
7.1 Natural-born essentialist 119
8. Conclusions

Chapter 4. *Mindreading and assumption of universality. Part 1)*
A primary form of representational mindreading system

1. Introduction 125
1.1 Being metarepresentational creatures 127
1.2 The nativist account does not imply full-fledged mindreading at birth 129
1.3 Against nativism 131
1.4 Nativism vs. constructivism on the propositional nature of attributed beliefs 132
2. False belief computation as TOM signature 135
2.1 The introduction of implicit false belief tasks and the new experimental paradigms 136
2.2 The problem of ' replicability' 137
2.3 Onishi and Baillargeon's «groundbreaking paper» 138
3. Taking into account other's perspective 141
3.1 VSP-taking and mentalization in infants: «susceptibility to others' beliefs» 142
3.2 *VSP-taking and mentalization in infants: visual access and consequent beliefs attribution in 18-month-olds* 146
3.3 Neural signatures in preverbal infants for representations attributed to others 148

3.4 Neural indication for very young infant appreciation of belief causal role in action interpretation 151

4. «Learning about the world through the lens of others» 153

5. Conclusions 157

Chapter 5. ***Mindreading and assumption of universality. Part 2) Propositional nature of attributed beliefs***

1. Problems about the nature of attributed beliefs 159

1.1 The weak sense of belief concept in infancy 160

1.2 Representational, propositional and conceptual dimensions for infant beliefs 162

1.3 To be believers infants do not need to be speakers 166

1.4 Davidson's triangulation: the belief notion revisited 167

1.5 Teleosemantic proposal 171

2. Almost like a belief: the minimal mindreading account 172

2.1 Song, Onishi, Baillargeon, and Fisher's experiment: updating without Encountering 176

3. Belief file cognitive structure underlying epistemic attribution 176

3.1 Two crucial values of belief file for learning about the world 178

3.2 First-person mental representations cannot vanish 179

4. Epistemic relations with object: remarks on Recanati's mental files 181

4.1 Mental files adapted for infant theory of mind: Perner and colleagues' Proposal 183

4.2. The crux of aspectuality 185

4.3. Belief attribution for dual identity objects 186

4.4. Aspectuality of belief representations in two helping tasks 188

4.5. Kampis' (2017) test of false belief about object identity 192

5. An unitary TOM observed from a neuroscience perspective 194

6. Attempts to solve the developmental gap from a cognitive point of view: the referential communicative bias 197

6.1 Helming, Strickland and Jacob's solution committed with the natural pedagogy theory 198

6.2 Grasping the relevant communicative referent 201

6.3 Westra and Carruthers' proposal: from referential bias to semantic Incompetence 201

7. Conclusions 202

Chapter 6. *Limits of natural pedagogy model*

1. Introduction: Epistemological side of natural pedagogy 207
1.1 Knowledge without justification. «The autonomy of knowledge with respect to justification» 209
1.2 Foundational and propositional nature of testimony 210
1.3 Moderate inferentialism and local reductionism for the credibility of the attester 211
1.4 The circularity of trustworthiness 213
2. Learning by trusting 214
2.1 Trustworthiness and reliability of the teacher (attester) The conquest of trust 216
2.1.1 Verifying reliable sources 217
2.1.2 Modalities of communication influence infants' evaluation of informants 219
2.1.3 Strengthening of monitoring capacities 220
2.2 Evaluating trustworthiness 222

3. What kind of pupils for natural pedagogy theory? 225
4. Low-level and high-level inferential process for triggering and sustaining natural pedagogy system 226
5. The theory of natural pedagogy under attack 227
5.1 Explicit teaching through ostensive cues: further remarks 230
5.2 Marion Vorms' critique 231
5.3 Nakao and Andrews' critique: around the notion of opacity 235
6 Natural Pedagogy and language according to Mattos and Hinzen 236
7. Limits of natural pedagogy 238
7.1 Natural pedagogy and deep certainties 239

Chapter 7. *Beyond the limits. Social biofeedback and natural pedagogy theories*

1. Introduction 243
1.1. The two assumptions of social biofeedback theory: lack of emotional introspection and the social construction of the infant's inner emotional Self 244

2. The Contingency detection device 246
3. Markedness of emotion expressions and its implications 248
4. Pretend component 250
5. What is shared and what is not by social biofeedback and natural pedagogy 253
6. Psychopathological implications? 256

Chapter 8. ***Conclusions*** 259

1. Epistemic trust as the basic assumptions of natural pedagogy theory 260
2. Early comprehension of the referential power of ostensive cues 261
3. The cognitive biases of the pedagogical stance 263
4. The supposed modalities of cooperation between natural pedagogy and Mindreading 265
5. Propositional nature of infant beliefs 267
6. Bounder of knowledge domain via natural pedagogy system 269
7. Values and limits of natural pedagogy 269

Acknowledgments 273

References 275

Introduction

1. *Eating a soup*

Imagine. A man enters a room in an experimental setting where a 12-month-old baby sits down on her mother's limb in front of a table. In the room there is a little kitchen that infant can see if she wants turning her head just a bit. When the man enters, he stops in front of the table and looks at the baby's eyes for a while keeping wide-eyed. He greets her making particular vocalizations, calling her name, and smiling while he keeps eye-contact with the child. Her name is Susi. Susi's attention is captured by the man's behaviour. The baby looks at him, who now turns his gaze toward the kitchen and points his arm and finger to same direction. The baby follows his gaze and his gesture, and the direction indicated by his finger. At this point he walks toward the kitchen, handles some pots, grasps a dish, and pours the soup into the dish, takes a spoon, and another plate with vegetables. Finally, he sits at the table in front of the baby. Keeping eye contact, he starts to have his lunch. He exaggerates his movements while he uses the spoon, and fixing constantly the baby, he manifests emotions of joy and pleasure, and says: "I like the soup! It's very good!". Besides the soup, on the table, there is another plate with vegetables. Now the man is pointing to this novel plate and he decides to taste the content, but vegetables provoke exaggerated expressions of disgust.

In our tale, the man performs several actions, and through these actions he is transferring a lot of practical information to Susi. His actions rely on two kinds of intentional attitudes, which can unfold simultaneously but more often subsequently, and remain independent each other. We may define the first kind of action and relative intention as communicative: the man wants to "tell" his audience that for some reason he wants (or desires, or has the goal) to go to the kitchen for … taking his lunch. He manifests such goal by referring to the kitchen through particular and structured gestures. His audience cannot immediately know why he is referring to that place or to something that there is in the kitchen. However, Susi understands that there is a relation between the man and the kitchen transmitted through eye-contact, gaze shift and pointing, namely, through what it is classified in the literature after Sperber and Wilson

(1986/1995) as *ostensive signals*. These signals seem to capture Susi's attention and she constantly looks at him while he is preparing the dish in the kitchen. While he is sitting in front of her, she observes how he handles that strange object, the spoon, that he is using for eating something he calls "soup", which is very good.

What is Susi thinking about?

According to natural pedagogy account, Susi's thought might be synthesised in this way: "The 'soup' is very good! While that green stuff in the other plate is bad!". In a few words, Susi interprets the communicative relation with the man as a teaching setting where the message to be learnt regards the referents of the ostensive signals, i.e. the 'soup' and the 'plate with vegetables'. Therefore, Susi is not immediately thinking that the man likes the soup and dislikes vegetables. Rather, the target objects of her thought are 'soup' and 'plate with vegetable'. In this way, the theory of natural pedagogy predicts that she acquires an important information which can be used and applied to further social interactions.

1.1 The man who wanted to teach how to eat a soup with the spoon

Born and immersed in human cultures, infants see and observe constantly so many actions that appear to be (and actually are) arbitrary, even counterproductive, but at the same time, they are relevant and necessary to be learnt in order to establish fundamental social exchange and promote a normal style of life. Therefore, the intergenerational transfer and the consequent cultural stabilisation of human technological habits, social conventions, folk traditions actually pose «a learnability problem for the purely observational learning mechanisms» (Csibra&Gergely 2011, p. 1151). Thus, the role of a benevolent teacher, intended as a guide who marks and emphasises the relevant aspects of these actions through specific communicative signals, is necessary.

In our tale, man's purpose was to teach Susi unfamiliar action that involves totally novel objects (probably) never used by the infant. He manifested such intention through a variety of communicative signals that serves to emphasise the sequence of structured and (instrumental) rational movements aimed to get the soup. What are the cognitive aspects relevant in such communicative-teaching relation? Does Susi interpret the two

(communicative and instrumental) actions depicted in the tale in the same way? Is there the possibility that the very sequence of the events influences the interpretation of the actions? The natural pedagogy theory, and further experiments connected to it, provide a model that has the ambition to answer to all these questions.

2. Natural pedagogy theory in a few words

The natural pedagogy theory proposed by György Gergely and Gergely Csibra (Csibra&Gergely 2006; 2009; 2011; Gergely&Csibra 2005; 2006; Gergely et al. 2007; Gergely 2007a; 2011) has the ambitious scope to provide a model of infant social learning. This particular social learning system is funded on a communicative triadic relation constituted by an adult, a child and a referent that can be concrete as an object (with its functions) or abstract like a word. In this sense natural pedagogy represents «a communicative system of mutual design specialized for the fast and efficient transfer of new and relevant cultural knowledge from knowledgeable to ignorant conspecifics» (Gergely&Csibra 2005, p. 463).

The theory focuses on «the transfer of relevant cultural knowledge through ostensive communicative […] manifestations» (Gergely&Csibra 2005, pp. 470-471). Therefore, we can define "pedagogy" as «a specialized human-specific cognitive system dedicated to cultural learning» (Gergely&Csibra 2005, p. 470) grounded on ostensive (verbal and non-verbal) communication expressed by eye-contact, smiles, particular vocalizations, infant-directed speech (IDS, or *motherese*), contingent reactivity, deictic gestures, joint attention. The domain of transferable knowledge can encompass: artifact functions and properties, novel means actions, first words, gestural symbols, social practices, rituals.

One of the revolutionary aspects of natural pedagogy consists, in my opinion, in representing infant cognition as a complex system devoting to assimilate knowledge from the social surrounding environment. In order to acquire information, evolutionary adaptation has construed an innate apparatus that makes infants sensitive to catch particular communicative signals that orient their attention and trigger a *referential expectation* in infants' mind. In this view, «preverbal human infants are prepared to receive culturally relevant knowledge from benevolent adults who are, in turn,

spontaneously inclined to provide it» (Csibra&Gergely 2011, p. 1154). Therefore, *ostensive signals can (a) allow the recognition of a potential teaching context and (b) sustain pragmatically the adults' intention to teach.*

The pedagogical stance construed by the adult-infant communicative relation triggers three biases in infants' mind, or in other terms, three different inferential processes (called 'assumptions') about the referential object of the informative transmission, and about the source of the information. I briefly anticipate them here but they will be presented and discussed in more details in Chapter 3.

1) The first assumption is the *epistemic trust*, that can be defined as the infant deferential attitude towards informative sources. The adult ostensive approach makes infant trustful and triggers the presumption of relevance about the new information transmitted, according to the *relevance theory* that will be one of the topic discussed in Chapter 1.

2) The second assumption (termed *assumption of generalizability*) predicts that ostensive-pedagogical manifestations facilitate infants «to convey information that is generalizable to the object class that the referent belongs to, considered as part of universally shared cultural knowledge about the object kind» (Gergely et al. 2007, p. 140).

3) The third assumption (*assumption of universality*) predicts that whatever the child learns in pedagogical context will be known by everyone. «If someone knows something, everyone knows it» (Csibra&Gergely 2006, p. 273). In other terms, it is presumed that what it is achieved in pedagogical condition is known and universally shared through a spontaneous epistemic attribution to the other members of the community, even if they are not present in the contingent teaching context. I assume that universality assumption is crucial for the efficacy of knowledge transmission and above all for the maintenance of cultural knowledge under the form of conventions and common ground beliefs between generations. Such implication of natural pedagogy would represent a fundamental strategy for cultural transmission enabling to optimize the reinforcement of the social practices, beliefs, and values of communities. Learning from others implies not only the acquisition of practical competences but also norms, rituals and beliefs of the own cultural group. Everywhere «children draw on a repertoire of cultural learning strategies that optimize their participation in and acquisition of the

particular practices, beliefs, and values of their community» (Legare&Harris 2016, p. 633). In this regard, Tomasello acknowledged that in virtue of natural pedagogy «human children do not just culturally learn useful instrumental activities and information, *they conform to the normative expectations of the cultural group* and even contribute themselves to the creation of such normative expectations» [my italics] (Tomasello 2016, p. 643).

The focus of my research is to investigate the nature of the epistemic attribution predicted by universality assumption. My hypothesis is that the omniscience bias is the result of a cooperation with a primary form of mindreading. For what concerns the relationship between natural pedagogy and mindreading faculties, I deem there are two unsolved questions:

i) Does natural pedagogy theory claim that a communicative intention is not part of theory of mind[1]?
ii) Does natural pedagogy theory claim that the ascription of informative contents to others can be not part of mindreading system[2]?

The main purpose of the present research is to answer these questions.

2.1. The supposed independence of natural pedagogy from mindreading

One of the most original contributes provided by natural pedagogy consists in the fact that «pedagogy offers a novel functional perspective to interpret a variety of early emerging triadic communicative interactions between adults and infants about novel objects they are jointly attending to» (Gergely et al. 2007, p. 139). Csibra, Gergely and cooperators claim that childhood attention (in terms of referential expectation) is

[1] A similar question has been arisen by Vorms (2012, p. 532) in a footnote of her paper that I will discuss in Chapter 6: «is natural pedagogy committed to the assumption that ascribing a communicative intention is not part of mindreading?». However, she doesn't employ any argumentation to answer.

[2] In my research I adopt the terms "theory of mind" and "mindreading" as interchangeable. In footnote 1 of Chapter 4 I explain the reason of such equivalence that is a motivated tendency recently emerged.

directed to action observed according to what is called «object-centered interpretation», rather than to the agent performing the action itself and her mental states (beliefs, desires, preferences) according to the «person-centered interpretation» (Gergely et al. 2007; Gergely 2007a). This fact is due to an important assumption at the base of the theory: the particular role given to the adult-infant interaction under an adaptive and cognitive point of view; a cognitive role that puts the pedagogical system in a primary, autonomous and independent status from mindreading within the *working-in-progress* architecture of infant mind. Indeed, Gergely and Csibra predict that pedagogical system is independent on mindreading system:

> the ability to teach and to learn from teaching is a primary, independent, and possibly phylogenetically even earlier adaptation than either language or the ability to attribute mental states. Having language and a theory of mind would no doubt assist both teaching and learning from teaching but […] they are not necessary prerequisites for pedagogical knowledge transmission. On the contrary, it seems to us equally possible that the cognitive mechanisms that had independently evolved to support pedagogy may have contributed to the subsequent evolution of language and theory of mind (Csibra&Gergely 2006, p. 250).

3. Aims of the present research

My thesis is an attempt to answer i) and ii) providing a model of early mindreading that is different from that one criticised by Gergely and Cisbra (2005; Csibra&Gergely 2006; 2009; 2011; Gergely et al. 2007; Gergely 2011). I suggest that natural pedagogy cooperates with mindreading which allows the ascription to others of informative contents having the form of beliefs, i.e. the form of propositional attitudes. I will argue my proposal in Chapter 4 and Chapter 5 where I will present several experiments about infant belief attributions (Kampis et al. 2013; Kampis 2017; Kovács 2015; Southgate&Vernetti 2014) conducted by cognitive psychologists belonging to (or very close to) Cognitive Science Department of Central European University in Budapest.

My purpose is to make evident the cooperation between natural pedagogy theory and infant primary form of mindreading, at least from two aspects. Firstly, the extreme

flexibility of belief attribution processing makes available the universality assumption for the benefit of natural pedagogy; then, the communicative dyadic interactions bias to some extent the comprehension of others' instructions in non-pedagogical contexts.

For what concerns the first horn of the puzzle, in Chapter 2, I will analyse two crucial aspects of infant ostensive communication:

1) *when* and *whether* infants understand that the very manifestation of a communicative (speech) act *refers to something*, i.e. whether the manifestation has primly a *communicative function* that can be «separable from its lexical content», just like those utterances, for example, that have the form of speech and that «can be interpreted as communicative even when the content cannot be comprehended» (Vouloumanos et al. 2014, p. 872; see also Martin et al. 2012; Vouloumanos et al. 2012).
2) *when* and *whether* infants understand to be the *addressees* of the communicative act.

The widespread literature about the early development of theory of mind attempts to explain. The infant capacity to *read*, to ascribe correctly preferences, desires, simple epistemic contents (like detecting the right location of objects) to other persons, animals, robots, geometric figures. However, a communicative intention is quite different from other kinds of intention because of the constrained way by which it is manifested. Csibra wondered whether in the case of communicative intention, it is adequate to speak about a cognitive process of recognition accomplished by early infants (Csibra 2010, p. 160). In fact, a decoding process of specific amodal behaviours probably is the principle actor able to affect infants from birth, as I will discuss in Chapter 1. In particular, in Chapter 2, I will present in brief the "fast-track modulator" model proposed by Senju and Johnson (2009), according to which infants early detect communicative signals normally provided through multimodal interactions as mutual eye contact and particular vocal intonations. In virtue of such detecting processes, infants can comprehend in some circumstances the whole referential power of ostensive signals.

In Chapter 6, I attempt to sign the limits of natural pedagogy and, at the same time, to clarify the misleading identification between pedagogical stance and contextual uses of ostensive signals. From an epistemology of testimony perspective, *natural pedagogy is supposed to be one of the main cognitive system responsible of cultural beliefs fixation*. This fact poses, in my opinion, natural pedagogy close to some Audi's positions (2015) and Wittgenstein's insights advances in *On Certainty*.

In Chapter 6, I will explain how the maturation of the full form of mindreading inhibits pedagogical instances. Mindreading in turn becomes necessary for social learning in more sophisticated context where children have to «differentiate trustworthy, benevolent, and reliable communicative sources of information from communicators who are unreliable, uninformed, or downright bad intentioned providers of useless or deceiving information» (Gergely et al. 2007, pp. 145-146). Therefore, when children increase the capacities to self-evaluate the source of information, and begin to hold introspective processes, they inhibit and block those parameters that allow the very trigger of pedagogical stance.

Finally, in Chapter 7, I reject a potential involvement of natural pedagogy in the theory of social biofeedback (Gergely&Watson 1999) as proposed by Gergely (2007b) first, and then by Gergely and Unoka (2008a; 2008b; 2008/2013). In their view, natural pedagogy would play a crucial role during the affective parental mirroring for the conceptualisation of emotions and the relative conscious inner access to emotional states. On the contrary, I suggest that during the affective dialogic relation the expressions of emotions conveyed through ostensive signals does not represent an explicit form of intentional teaching. This feature, instead, represents one of the inalienable element for the pedagogical relationship.

CHAPTER 1.

Action interpretation criteria for infants' mind

1. Communicative and instrumental acts

In the beginning tale, at first sight, it could seem that the baby is attuned with the intentions underlying the two agent's actions (the communicative and the goal-oriented action). Differentiating the two kinds of action is the first step for the comprehension of the cognitive mechanisms underlying infant action interpretation. Is the infant able to attribute to the agent the two different kinds of intentions? The communicative action is performed through few amodal, clear, transcultural gestures as ostensive signals are. Children are very sensitive to them from birth (Cecchini et al. 2011; Farroni et al. 2003; Grossmann et al. 2008), and the continuous cues (and habituation to them) may conduce infants very early to recognise communicative context. Csibra (2010) suggested that we don't need to involve any mindreading capacity in order to understand these communicative signals and their referential nature in early stages of development. This will be the topic of Chapter 2.

What about, instead, the goal-oriented (instrumental) actions of laying the table and eating the soup with a spoon? According to several researchers (e.g., Woodward (1998), Apperly and Butterfill (2009)), goal-oriented actions imply mindreading capacities (or a sort of "minimal" kind of theory of mind[1]) in order to be understood by human infant interpreters. However, in the last fifteen years, intriguing experiments suggest the great ability shown by very young infants to understand goal-oriented action in virtue of recognition (and thus attribution) of efficiency to the action itself (Csibra&Gergely 1998; 2007; Csibra et al. 2003; Csibra 2003). Last but not least, according to the recent Pomiechowska and Csibra's (2017) study also the instrumental actions can be interpreted in a different way if they are preceded by communicative cues. Therefore, the question is: do infants read intentions or actions first?

[1] I will discuss it in Chapter 5 (§2).

2. *Reading actions*

«An agent's intentional state is about something in the world» (Carey 2009, p. 158). *Desiring*, *wanting* something are paradigmatic examples of intentional states. At the same time, also *referring* can be considered as another *intentional relation* (Carey 2009, p. 159). In any case, we must acknowledge, following Susan Carey, that ascribing intentions means attributing *propositional attitudes* to agents (Carey 2009, p. 159)[2]. But have infants access to persons' internal mental states? And precisely, are human infants able to metarepresent other's mental states in the form of propositional attitudes, as argued for example by Alan Leslie (2000)? In this regard, do infants share such core mentalistic characteristics with adults (Carruthers 2013; Tirassa et al. 2006)? Or do they undergo a conceptual change during developmental phases?[3] (Butterfill&Apperly, 2013; Carey&Johnson 2000). The best way to provide a response to these questions is looking at the cognitive ability under the lens of pragmatics with particular reference to "action-oriented paradigm".

2.1 "Action-oriented paradigm" in infant cognition

In 2013, Engels and colleagues recognised a pragmatic turn in cognitive science, and the consequent achievement of the "action-oriented paradigm" (Engels et al. 2013, p. 202). The fact that the cognitive processes are connected with action emerges clearly also from the long phenomenological tradition (see Dominey et al. 2016 for a review[4]).

[2] The propositional nature of belief attributions will be discussed in Chapter 4 and Chapter 5.

[3] In the fourth, and above all, in the fifth chapter I will tackle in more details this topic and, on the basis of the most recent findings, I will argue that infants share core mentalistic characteristics with adults and do not undergo a conceptual change during development. Moreover, I will suggest the propositional nature of infant attributed beliefs.

[4] The authors point out how action-oriented paradigm has influenced cognitive sciences and their domains as social cognition, perception, sensorimotor entrainment, language acquisition and so on), and also how such paradigm conditioned neuroscience studies («with the notion that brains do not passively build models, but instead support the guidance of action»). Furthermore, the authors employ not only the potential application of enactive control principles to action and perception in robotics, but also the

Varela, Thompson and Rosch defined cognition as "embodied action" (Varela et al. 1991). In this framework «cognition should be understood as the capacity of generating structure by action» (Engels et al. 2013, p. 202). In other terms, we can assert that *cognition is a form of practice*. Therefore, 'action' is the first key-term that we meet in this research-pathway. Action is not a synonymous of movement, its notion

> «implies that actions (i) are driven by goals and that they can reach these goals or fail to do so; (ii) often involve some degree of volitional control; (iii) require planning and decisions among alternatives; (iv) involve prediction or anticipation of an intended outcome; (v) are often, albeit not always, associated with a sense of agency, that is, the agent's conscious awareness of carrying out the particular action and of its goals» (Engels et al. 2013, p. 203).

The abovementioned action-features involve an agent who acts. In other terms the intentionality of an action concerns its origin from the point of view of the actor who, according to what Donald Davidson said, performs his act «in the light of his beliefs and desires» (Davidson 2001, p. 84).

The general topic of my research is not focused on the phase of individual action production, rather on the phase of *action interpretation* made by infants who see and are engaged in actions displayed in the social environment. Action is generated by complex, conscious and sub-personal mental states; therefore, the problem is to discriminate whether infants are interested (and able) to understand the agent's volition – where 'volition' is to be intended in the terms provided by Wilfrid Sellars (deVries 2016)[5]. The action performed that will be analysed concerns a kind of *teaching* act. Therefore, I assume that this specific act is the result of a clear volition of teaching, i.e., the result of an intentional communicative act performed by the adult (teacher) with the aim to transfer relevant information about the surrounding world to naïve learners.

The arising dilemma regards the following interconnected questions:

- In order to comprehend others' actions, what cognitive means do infants use?

---

practical applications for what concerns education, developmental and psychopathological disorders , the development of neural prostheses (Dominey at al. 2016, p. 333).

[5] As Willem deVries (2016) summarises Sellars' thought: «An intention is a thought that motivates one to realize its content» (Stanford Encyclopedia, https://plato.stanford.edu/entries/sellars/#5.1).

- Do infants have to understand the agent's intentions? Or should they understand the action itself (i.e. its structure and its perceptual/kinesthetic features) without particular capacities of metarepresenting others' mental states?

It doesn't matter for our purpose to investigate the origins of intention in action. In this regard, we can assume different positions without solving the puzzle about the interpretation of an action made by an agent. We could agree with McDowell who claimed (following Sellars' suggestions) that «the proper form of expression for intentions in action is 'I am doing such and-such'» (McDowell 2010, p. 417), on the one hand. On the other hand, we can accept Davidson's account, according to which «someone who acts with a certain intention […] has something in mind that he wants to promote or accomplish» (Davidson 2001, p. 83)[6]. In both cases, the options to verify are
i) whether infants really need to 'read', i.e. to represent in some way these expressions in order to understand the action they are looking at;
ii) whether they don't need at all;
iii) whether they need, at least, to read only some kinds of intention in early stages of their cognitive development.

Part of the long debate in developmental psychology about contingent responses in infants concerns *where* (and under *what conditions*) the main attentive focus is addressed by the infants: either towards the person who acts, or rather towards the actions (to intend also as specific and structured gestures) accomplished[7]. So far, a good explicative strategy to read correctly the capacity and the target of childhood action interpretation has been to measure the infants' predictive power addressed to others. Indeed, "action interpretation" entails also "action prediction" in terms of expectations. This fact represents one of the most important aspects that allows making insights about

[6] John McDowell (2010) argued that these different forms used to intend intention in action are opposite.

[7] Everyone shall agree with Gergely when he claims that the researches and the different accounts about «the early development of understanding intentional action and agency» started 40 years ago together with the «research interest in the ontogenetic origins of understanding and representing other minds» (Gergely 2011, p. 76), inaugurated by the seminal studies of Premack and Woodruff (1978), and Wimmer and Perner (1983).

infant cognition. Through the experiments that will be illustrated, the attempt is to provide a correct reading about such predictive attitude shown by infants during opportune tasks.

3. Types of action and types of agent: a useful distinction

Information received by the infants are transmitted through actions. From the birth newborns watch actions. An action, thus, it is bound to something recognisable to the infants in the environment, in order to be understood and eventually reproduced appropriately. As Vorms (2012) pointed out too, the information being understood by infants (or better, in pedagogical terms, generalised and universalised by infants) is about specific actions shown and related to object(s); in particular objects involved in actions determined by agent(s). From here the importance of understanding the relation between agent(s) and object(s), or in other terms, the agent's intentional state.

Csibra (2003; 2010) and Gergely (2011) proposed a criterion for infant understanding of intentional agency, according to which infant's point of view must be analysed in respect of the actions, i.e. the perspective of observation by which the infant watches the action and the relative degree of her/his involvement. In other words, *action interpretation may be influenced by the role played by the infant*: the child may limit to observe the action without any commitment (third person's view), or she/he may be involved by the performer with communicative acts (second person's view) (Jacob&Gergely 2012).

Going back to our example, when the man enters the room and attempts to catch the child's attention with ostensive signals and indicates the kitchen, and then the plates, the infant is observing the actions in second person perspective, interacting with the agent, and looking at the ostensive signals addressed with particular attention. The way of decoding such communicative signals determines a different inferential process characterised by specific biases (Csibra 2003; 2010; Gergely 2011). Therefore, distinguishing *instrumental action* from *communicative action* may be very useful if our purpose is to investigate the cognitive processes adopted by the child to interpret the sense, the goal and intention of the action. Recent researches induce to think that this distinction has neuronal correlates (Southgate&Begus 2013; Pomiechowska&

Csibra 2017). As for our tale, in this respect, a crucial fact that has to be underlined concerns the sequence of events. As I will try to show below, following Pomiechowska and Csibra's experiment (2017), the man's communicative referential action triggers biases which may exchange the interpretation of his following instrumental action.
Let's see more carefully the relevance of such distinction starting from the instrumental action interpretation.

4. Instrumental action (interpretation)

«An action is goal directed if it is performed not for itself but to achieve an end; in other words, if it is an instrumental action» (Csibra 2003, 448). This is a definition of a goal-oriented action, but it needs to be more precise if we have to understand how this action is interpreted.

> «Interpreting something as an instrumental action that has been performed to achieve, and gets its meaning from, a particular goal state requires consideration of many things – not just the action and its end state but also the environment in which it occurs» (Csibra 2003, 448).

Using the violation-of-expectation looking time paradigm (VOE)[8] Gergely and Csibra (2003; Csibra et al. 2003) challenged the current mentalistic stance apparently shown by infants since twelve-month-old to interpret goal directedness. Such mentalistic approach relatively to children cognition (Kelemen 1999) is divided into a "modularist" account[9] (Premack 1990; Baron-Cohen 1994; Leslie 1994; Premack&Premack 1995) and a "simulationist"[10] account (Meltzoff 2002). According

[8] See chapter 4 for more explicative details about such experimental paradigm.

[9] For *modularist* theories intentional mental states (like desire) are attributed to an actor thanks to an innate stimulus cue~~s~~ (such as gaze direction~~s~~) that activates a prewired mechanism whose computational output identifies the goal state.

[10] For *simulationist* theories the other's mental states are identified through a process of identification. Infants put themselves into the actor's shoes and simulate her intentional mental states as if they acted like the actor. Such infant subjective mental states are, thus, introspectively available and then attributed to the actor's mind.

to a general mentalistic view, if infants saw on a screen, for example, a small yellow circle jumping an obstacle and touching a large red circle, they would attribute «to the small circle a desire to get to the large circle and a belief about the impenetrability of the obstacle» (Gergely&Csibra 2003, p. 288).

Gergely and Csibra proposed alternatively that infants «can represent, explain and predict goal directed actions by applying […] the "teleological stance"» as described by the "naïve theory of rational action" (Gergely&Csibra 2003, p. 289; Csibra&Gergely 1998). The "teleological stance" «represents an interpretational strategy that seeks to construe an event in terms of goals» (Csibra 2003, p. 448). As Csibra acknowledges, it is akin to *intentional stance* descripted by the philosopher Daniel Dennett (1987), in virtue of two common features:

a) it is not «an explicit inferential system but a *bias* [my italics]» (Csibra 2003, p. 448);

b) there is an application of *rationality principle* ascribed by the infants to agent's mental state (in the case of mentalistic stance), and to actions itself in terms of "efficiency" (in the case of teleological stance).

Therefore, the two stances are different because the teleological stance «does not attribute [rational] mental states to the agents» (Csibra 2003, p. 448). Teleological stance provides an explanatory relation oriented to the scopes and not to the causes among three elements of observed and future reality: «the *action*, the (future) *goal state*, and the current *situational constraints*» (Gergely&Csibra 2003, p. 289). In short, teleological stance entails:

1) representation of the goal,

2) representation of the physical constraints present in the action's scenario, and

3) representation of the means for achieving the goal (given the environmental constraints).

Such representations of the infant are supported by the "principle of efficiency" strongly bound to the features of the action itself. As Susan Carey nicely summarised, there is not «any relation between an agent and a desired state explicitly represented. Rather, it is the action itself that is represented as goal-directed» (Carey 2009, p. 166).

4.1 Teleological stance and mindreading overlapping

«The teleological properties of actions» are captured by infants «*without* presupposing the availability or necessarily relying on the *contents* of the agent's mental representations of reality» (Gergely 2011, p. 84). In contrast with mentalistic stance, «the rationality principle is always applied to the contents that the actor's mental states represent» (Gergely&Csibra 2003, p. 289) rather than to representations of desire and belief themselves. Applying the rationality principle in teleological stance means to evaluate the efficiency of the observed action. In other terms, the teleological evaluation judges the means necessary for achieving the goal as efficient. «In teleological interpretations, judgments about the rationality of means always translate into judgments about 'efficacy'» (Gergely&Csibra 2003, p. 290).

> «This teleological evaluation of the efficiency of means should provide the same results as the application of the mentalistic stance as long as the actor's action is driven by true beliefs» (Gergely&Csibra 2003, p. 290).

According to Gergely and Csibra, the two stances are independent and the teleological one is more ancestral and it matures before. When mindreading system starts to work, «representational contents of the agent's epistemic mind states» (Gergely 2011, p. 84) will be evaluated by infants and the teleological stance will be transformed into a mentalistic intentional stance, according to the following schema illustrated by Gergely and Csibra (2003, p. 289; Gergely 2011, p. 85).

Therefore, we will have that while «teleological representations provide explanations and predictions for observed actions by relating three aspects of reality via the rationality principle, […] mentalistic action representations involve three types of intentional mental states whose contents correspond to the representational elements of teleological action representations» (Gergely 2011, p. 85). Even though the two systems are independent, and one ontogenetically precedes the other one, they overlap at a certain stage of normal cognitive development.

Initially, Gergely and Király (2003) proposed that this overlap takes place between 14 and 18 months of age through a «shift from a teleological construal to a mentalistic interpretation» (Gergely 2003, p. 119). On the bases of imitation tasks

construed following Meltzoff's experiment (1988)[11], they supposed that comparing to 14 month-old infants, 18 month-olds were more sensitive to communicative-referential behavioural cues, and this fact led infants to interpret the demonstrator mentalistically, ascribing to him communicative intentions.

> «Communicative cues come to be interpreted by the 18-month-olds mentalistically in terms of the intentional stance as signalling and setting the frame for a subsequent communicative act that will involve the demonstration of some new and culturally relevant information with the intention on the demonstrator's part to make it available for the infant to learn about it» (Gergely 2003, p. 127).

A few years after the publication of this essay, when Gergely himself and Csibra proposed the natural pedagogy theory, they showed more wariness to involve mindreading skills in early infant ascription of communicative intention to agents in social learning context. I will discuss in the following chapter further researches, many of them conducted by Csibra himself, on precocious infantile sensitivity to interpret ostensive signals. The amount of collected data about such infants' cognitive capacity may suggest that there is no necessary entailment with mindreading capacity. Therefore, another kind of interpretation must be provided for explaining Gergely and Király (2003) results on infants with normal cognitive development. Natural pedagogy represents an alternative explanatory strategy, as we will see in Chapter 3.

4.1.1 The independence of the two interpretational schema in psychopathology

One of the best example to show the independent origins of teleology and mindreading comes from those psychopathological researches (see Fonagy et al. 2002 for a review) that investigated the «inhibited or distorted development of mentalization skills» (Gergely 2003, p. 119). In such cases, individuals suffer a regression to the level of non-mentalistic of social reality, «interpreting interpersonal events in terms of concrete, visible, physical outcomes of actions instead of being able to correctly infer the mental states driving the actions of their interactive partners or, in fact, of themselves» (Gergely 2003, p. 119). In other terms, the pathological subjects suffering a

[11] Meltzoff's experiment (1988) is descripted and discussed at the beginning of Chapter 3.

history of dysfunctional attachment apply teleological strategies to social relations. Therefore, the impairment of mentalistic capacities shown in personality disorders (Fonagy&Target 1997; Gergely 2002) may provoke the persistence of teleological interpretative strategies[12].

4.2 Naïve theory of rational action against motor mirroring mechanism for action understanding

Two conflicting positions animate the pluri-decennial debate around infant instrumental agency interpretation. On the one hand, there are those who assert that motor mirroring mechanism sustains human action understanding, enabling observers to understand other individuals' actions (for a review Sinigaglia&Rizzolatti 2011)[13]; on the other hand, others suggest that motor mirroring «is a corollary of action interpretation» (Pomiechowska&Csibra 2017, p. 1; Csibra 2007a; Hickock 2013; Prinz 2006; Hamilton 2013; Sartori et al. 2013)[14]. The extreme opposition of these two accounts is more evident in instrumental actions.

According to the first theoretical position, «goal attribution is achieved by the observer through simulation […] without engaging in any inferential processes» (Pomiechowska&Csibra 2017, p. 1). Hunnius and Bekkering (2014) claim in this regard that: «When infants […] observe […] actions in others, they can use their motor system to predict the outcome of the ongoing actions. [And they] come to an understanding of others' actions through the repeated observation of actions and the effects associated

[12] For more details about this issue, combined with the notion of normal and abnormal parent-infant dyad relation, see also Fonagy et al. (2007) and Gergely&Unoka (2008). For a review see in italian Marraffa&Meini (2016).

[13] Sinigaglia&Rizzolatti (2011) provide a general account of the mirror mechanism, and for more up-to-date researches see the volume 369, issue 1644 of *Philosophical Transaction of the Royal Society B*, published in 2014, and dedicated to the theme "Mirror neurons: fundamental discoveries, theoretical perspectives and clinical implications", edited by Francesco Ferreri e Giacomo Rizzolatti. I point out in particular the review article by Rizzolatti&Fogassi (2014), Woodward&Gerson (2014), Hunnius&Bekkering (2014), Simpson et al. (2014), Marshall&Meltzoff (2014), and Vivanti&Rogers (2014) for the relation between autism and mirror neurons in learning and in teaching.

[14] For a general account on mind, perception and agency interpretation characterised by «top-down expectation or prediction» see Clark (2013).

with them» (p. 1). The background assumptions of such account consist in the fact that «active action experience is crucial for infants' developing action understanding», and «infants have *plenty of opportunities* [my italics] to form associations between observed events and learn about statistical regularities of others' behaviours» (Hunnius&Bekkering 2014, p. 1).

A rich first person view is implicated in this perspective, as Woodward and Gerson (2014, p. 1) well highlighted: «infants' own goal-directed actions influence their analysis of others' goals. […] [C]ognitive systems that drive infants' own actions contribute to their analysis of goals in others' actions». According to the direct-matching hypothesis provided by Rizzolatti and colleagues, «an action is understood when its observation causes the motor system of the observer to 'resonate'» (Rizzolatti et al. 2001, p. 661). Such resonance allows the observer to understand the outcome, and thus finally, the goal of the action thanks to the fact the observer knows «its outcomes when he does it» (Gallese et al. 2004, p. 396). In other words, as Csibra summarised, «action mirroring provides a simulation device for goal understanding by automatically and mandatorily duplicating the observed action in the observer's motor system» (Csibra 2007, p. 436)[15].

Before proceeding, I need to advance a little but relevant clarification that circumscribes the "endless" experimental literature about MNs. Here, we are speaking about instrumental actions, i.e., sophisticated actions that involves objects too. Therefore, it is necessary to leave apart all those significant and intriguing experiments conducted on imitation of facial emotional expression performed by infants and children. Like for example the case of Rayson and colleagues' study (2016) which showed «movement simulation during observation of facial expressions» in 30-month-old children «as they observed videos of dynamic emotional and non-emotional facial expressions». These results have been obtained by investigating, through electroencephalography (EEG)[16], the phenomenon of mu rhythm desynchronization[17],

[15] Csibra (2007, p. 436) proposed that «the primary function of action mirroring is not action understanding in terms of goals but predictive action monitoring. [He] suggest[s] that action mirroring in the observer is achieved not by direct matching but by emulative action reconstruction».

[16] EEG-imaging methods are very promising for studying social-action representation in infancy and early childhood, as Liao et al.'s (2015) study indicates too (See also footnote 2 in Chapter 2 about ERP (Event Related Potential) technique).

that is a significant index of MNs activity used by Pomiechowska and Csibra (2017) too for instrumental actions, as we will see below.

An important and controversial point of simulation account based on mirror neurons mechanism concerns, in fact, not only the hierarchal role of motor system activation on infant behavioural response, but also the assumption of a rich and structured first person who comprehends other's agency starting from the self-perceptive experience. A strong first person authority in infants presumes an innate capability to access to the own inner world, own mental states and the faculty to ascribe them to others. However, a more parsimonious explicative argumentation is possible. It is provided by "the naïve theory of rational action" that indicates how infants have an innate capacity to ascribe efficiency to action without any inferences on the own and the agent's mental states. The inferences regard the structure of the action and the environmental elements following a top-down process that modulates and influences motor activation system (Csibra 2003; 2007; 2010; Pomiechowska&Csibra 2017).

4.2.1 Southgate and colleagues' experiments in favour of the application of the rationality principle in instrumental action interpretation

In the last ten years, Victoria Southgate and colleagues provided the most decisive contributions to the naïve theory of rational action, sustaining also the proposal that instrumental action interpretation is not driven directly by motor activation. Southgate and colleagues analysed, for example, the application of efficiency principle in 6-8 month-old infants who observed biomechanical impossible events (Southgate et al. 2008), measuring the number of action steps the agent (a human arm showed in a videotape) performed for achieving its goal. The assumption was that the less number of steps it took to achieve a goal, the more efficient was the action. The aim was to test whether infants recognised «actions as goal-directed on the basis of their experience […], or whether their perception of a goal-directed action [was] based on the

[17] Sensorimotor cortex is activated during action execution and observation both in adulthood and in childhood, as it has been discovered measuring alpha suppression , or mu rhythm, whose oscillations frequency are around 8-13 Hz in adults and 6-9 Hz in infants (see e.g., Pineda 2005; Pineda&Hecht 2009).

recognition of a specific event structure» (Southgate et al. 2008, p. 1059). The researchers showed how infants extended goal attribution also to a human arm that reached an object overcoming obstacles with snake-movements. They concluded «that notion of goal is unlikely to be derived from infants' experience» (Southgate et al. 2008, p. 1059). If actions were understood for the fact that young observers directly mirror «the observed action onto their own motor system» (Southgate et al. 2008, p. 1060; Rizzolatti&Craighero 2004), triggering a motor simulation, it would follow that «only action that observers can themselves perform w[ould] be able to be simulated» (Southgate et al. 2008, p. 1060; Rizzolatti et al. 2001). Southgate and colleagues' experiment contradicted also Woodward's (1998) conclusions, according to which only «through experience with particular actions […], infants gradually come to construe actions in terms of attaining goals» (Southgate et al. 2008, p. 1060).

Four years later, Victoria Southgate and Mikolaj Hernik (2012) demonstrated that 9 month-old infants comprehend the structure of an oriented-goal action independently of the agent's preferences (Hernik&Southgate 2012), falsifying what Luo and Baillargeon claimed (2005) in this regard. Indeed, Hernik and Southgate (2012) showed that 9-month-olds expected the agent (represented by a red cube) to continue acting towards the previous goal (reaching a blue cylinder behind a barrier) even if additional choice-options (another brown cylinder) became available[18]. The results indicated that «there was no preference-related evidence. [Therefore, they] concluded that infants do not need to know about the agent's preferences in order to form expectations about its goal-directed actions» (Hernik&Southgate 2012, p. 714).

Scott and Baillargeon (2013) tested the application of rationality principle too, but without involving infrequent, impossible or odd actions.

> «In two experiments, 16 month-olds watched events in which an agent faced two identical goal objects; although both objects could be reached by typical, everyday actions, one object was physically (Experiment 1) or mentally (Experiment 2) more

[18] The use of objects instead of human agents may further indicate that there is no possible identification and simulation processing. This theoretical and methodological approach goes against also Woodward's (1998) interpretation according to which young infants can ascribe goals only to those agents who appear perceptually very similar to their own bodily aspect, and whose body-movements can be mapped onto infants' own motor repertoire (see also Jacob&Gergely 2012 for a critique against Woodward's account).

> accessible than the other. In both experiments, infants expected the agent to select the more-accessible object. These results provide new evidence that infants possess a general and robust expectation of efficiency» (Scott&Baillargeon 2013, p. 466).

Recently, also Liu and Spelke (2017, p. 35) accepted the abovementioned evidence, according to which very young infants expect agents «to move directly to their goals when no obstacles block their paths», or «on the least costly path that the environment affords». On these bases, they tested sixty 6 month-olds across three experiments with novel, curvilinear action trajectory, and they found that «infants expected minimally costly action when presented with a novel constraint, and extended this expectation to agents who had previously acted inefficiently» (Liu&Spelke 2017, p. 35).

4.3 Neuronal correlates for teleological reasoning

Southgate and colleagues (2014) investigated also whether in adults and infants there were correspondences between cerebral areas involved in instrumental action understanding. Hamilton (2006) indicated the inferior frontoparietal cortex as the brain area more involved in goal representation. Southgate et al. (2014) investigated whether the same areas were recruited in the infant processing of goal-direct event, using «a repetition suppression (RS) design, similar to that used with adults» (Southgate et al. 2014, p. 295). They found «a strikingly similar response pattern and location of activity as had been reported in adults» (Southgate et al. 2014, p. 294).

Last but not least, Southgate and Begus (2013) demonstrated, using neural indication sensorimotor-cortex activation measured by EEG, «that 9-month-old infants recruit their motor system whenever a context suggests an impending action, but that this recruitment is not dependent on being able to match the observed action with a corresponding motor representation». Their data support «the view in which motor activation is the result of, rather than the cause of, goal identification» (Southgate&Begus 2013, p. 828). This last finding is congruent with the recent Pomiechowska and Csibra's study (2017), that provides empirical foundations to the useful theoretical distinction operated by Csibra (2003; 2010) and Gergely (2011) between instrumental and communicative action interpretation.

4.4 Neural correlates for referential expectation in communicative actions: the Pomiechowska and Csibra's proposal

A lot of findings show the activation of motor system in action processing, suggesting the hypothesis whereby without simulating observed movements in the motor system, individuals should have difficulty in interpreting observed actions. Vannuscorps and Caramazza (2016) challenged such hypothesis across eight sensitive experiments in which «individuals born with absent or severely shortened upper limbs (upper limb dysplasia), despite some variability, could perceive, anticipate, predict, comprehend, and memorize upper limb actions, which they cannot simulate, as efficiently as typically developed participants» (Vannuscorps&Caramazza 2016, p. 89). The authors point out that their results are based on the same experimental materials and procedures used in those studies that have been interpreted as the best evidence in favour of motor simulation theories (Bosbach et al. 2005; Wilson et al. 2010). Therefore, future research should investigate deeply how (visuo-)perceptual and cognitive system encode information of body part movements, and in general, support interpretation of actions. A recent attempt is constituted by Becchio et. al.'s study (2018, p. 67) that reframe «the problem of direct perception in terms of establishing a […] measurable relationship between movement features and perceived» intentional states tied to goal oriented actions. Alternatively, it has been proposed that the activation of motor system «might not necessarily be due to the fact that motor mirroring provides bottom-up support for action understanding. Rather, mirroring might be generated in a top-down manner as a result of action interpretation» (Pomiechowska&Csibra 2017, p. 2).

> [T]he mirror-based account of action understanding proposes that the recruitment of the observer's motor system is a precondition of action understanding, while the action reconstruction and the social responding accounts argue that motor mirroring is one of the outcomes of action understanding (Pomiechowska&Csibra 2017, p. 17).

Accepting the theoretical distinction between instrumental and communicative actions (Csibra 2003; 2010; Gergely 2011; Jacob&Gergely 2012), the authors compared

«the levels of motor activity across instrumental and communicative actions», assuming that «the selected interpretation should directly modulate the level of motor activation» (Pomiechowska&Csibra 2017, p. 6). Therefore, they expected a different degree of activation depending on the kind of action observed.

> «To assess motor activation, [they] measured the mu rhythm (also known as the resting state sensorimotor alpha rhythm, 8-12 Hz) through electroencephalography (EEG). Attenuation (or suppression) of this rhythm is typically observed while participants execute movements or observe biological movements executed by others […] and is considered to reflect the activation of the sensorimotor cortex» (Pomiechowska&Csibra 2017, p. 7).

However, the authors were conscious that it was not enough to measure different levels of mu suppression for indicating a stronger action interpretation, because the instrumental action of grasping and reaching an object elicits a major quantity of neural population in comparison with the referential action of pointing. For this reason, the participants at the experiments «were presented with speech (*Look!*)[19] or a matched pure tone prior to the onset of the action» (Pomiechowska&Csibra 2017, p. 8)[20]. The presence of many ostensive signals would clearly «bias participants to suppress the instrumental interpretation of the observed acts» (p. 8) and foster them to construe the action as referential.

> «Under this assumption, identical grasping actions should be interpreted differently depending on the preceding sound stimuli, thus affecting levels of motor activation. Specifically, if the presence of speech changes the interpretation of grasping from instrumental to referential, less mu suppression should be recorded in the presence of speech than in the presence of a pure tone» (Pomiechowska&Csibra 2017, p. 8).

[19] «"Nézd csak!" ("Look!") recorded by a female Hungarian native speaker» (Pomiechowska&Csibra 2017, p. 10).

[20] Thirty-four healthy volunteers participated in the two experiments (mean age: 22 years, range: 18 to 27 years).

In video clips a human hand performs one of the following three object-directed actions: «(1) grasping an object, (2) reaching for an object without grasping it, and (3) pointing to an object with the extended index finger» (Pomiechowska&Csibra 2017, p. 9).

The study combines communicative and non-communicative sounds (pure tone vs. speech) during the different kinds of actions. In a particular experimental condition (Experiment 1), «one type of action was paired with one type of sound (pure tone – grasping, pure tone – reaching, pure tone – pointing, speech – grasping, speech – reaching, speech – pointing)». While in another experimental condition (Experiment 2) grasping, reaching, and pointing actions were paired only by sound (Pomiechowska&Csibra 2017, p. 11). The results indicate that:

> «When the context suggested a referential interpretation of the observed action, either due to the semantics of the witnessed gestures (i.e., pointing) or to the presence of speech, there was no sign of significant motor activation. This pattern of results suggests that *action interpretation is not dependent on the observer's motor system* [my italics] and that the presence of subsequent motor activation is conditioned by this interpretation: only conceiving of an action as instrumental, but not as referential, leads to the recruitment of sensorimotor cortices during action observation» (Pomiechowska&Csibra 2017, p. 16).

Another very important result of this experimental study consists in the fact that the presence of communicative speech signal before the onset of the instrumental action of grasping modifies the interpretation of the observed action. «The communicative context induced the expectation for referential signals rather than for an instrumental action» (Pomiechowska&Csibra 2017, p. 16), as predicted by Csibra (2003) and Gergley (2011). In this way, the object exchanges its role: from the *target* of an instrumental action, it becomes the *referent* of a communicative act.

Therefore, according to these results, we are authorised to suppose that only the interpretation of instrumental action is subordinated to motor activation. Following Vannuscorps and Caramazza's (2016) findings on patients with limb dysplasia, and Southgate and Begus' (2013) experiment, Pomiechowska and Csibra suggest that goal-

oriented action motor activation seems to be «the result of, rather than a contributor to, goal identification» (Pomiechowska&Csibra 2017, p. 20).

5. Context and neural mechanism for action interpretation

Even if we assume that action interpretation is not driven by action-simulation, the question about which neural substrates underlying top-down processes in action interpretation remains open. We must remember also that both instrumental and communicative acts are oriented towards an object. Therefore, it is reasonable to wonder if also object representation is influenced by top-down action interpretation itself. According to Yoon and colleagues (Yoon et al. 2008) preverbal infants remember the location of a toy better in an instrumental context than in a referential context, in which they tend to store the information about its appearance. Yoon and colleagues, indeed, tested

> «whether 9-month-old human infants can distinguish between communicative and non-communicative social context and whether they retain qualitatively different information about novel objects in these contexts. [They] found that in a communicative context, infants devoted their limited memory resources to encoding the identity of novel objects at the expense of encoding their location, which is preferentially retained in noncommunicative contexts» (Yoon et al. 2008, p. 13690).

The role played by the (recognition of) the context seems to be crucial for the natural pedagogy theory. As shown by Wurm and Schubotz (2016), contextual information modulates action recognition at different levels of processing.

> «We measured recognition performance of hardly identifiable actions that took place in compatible, incompatible, and neutral contextual settings. Our findings demonstrate that contextual information is effectively exploited during action observation, in particular when visual information about the action itself is sparse» (Wurm&Schubotz 2016, p. 1).

The recognition and the use of contextual inputs are the matter of the following paragraphs which will tackle the other horn of action interpretation, the communicative aspect.

6. Communicative context: the heritage of relevance theory for natural pedagogy

> «The notion of context depends on the notion of common ground. The common ground between two agents consists of mutual knowledge, mutual beliefs, and mutual suppositions that they can share. Examples are the social norms shared by the members of a particular community» (Bosco et al. 2004, p. 468).

As I will suggest in the course of the present dissertation, the theory of natural pedagogy has the ambition to explain the cognitive strategies that allow the transmission, the establishment and the sharing of such mutual (normative) knowledge. The main theoretical framework on which natural pedagogy is based is the *relevance theory* provided by Sperber and Wilson (1986/1995). Natural pedagogy model takes nourishment from the relationship both between relevance and human cognition, and between relevance and communication. Let's see in brief the former.

People *expect* to find something relevant, and they attempt to *select* a context in which that expectation is sustained, and through which they can maximise relevance (Bosco et al. 2004, p. 468). According to the relevance theory, an individual automatically tends to maximise the relevance of the inputs that he processes: «it is a matter of making the most efficient use of the available processing resources» (Wilson&Sperber 2002, p. 610). From here the "cognitive principle of relevance", according to which «human cognition tends to be geared to the maximization of relevance» (Wilson&Sperber 2002, p. 610). Relevance is described as «a potential property not only of utterances and other observable phenomena, but of thoughts, memories, and conclusions of inferences» (Wilson&Sperber 2002, p. 608).

> «Something is relevant to an individual when it connects with background information he has available to yield conclusions that matter to him […]. According to relevance theory, *an input is relevant to an individual when its*

> *processing in a context of available assumptions yields a positive cognitive effect*. A positive cognitive effect is a worthwhile difference to the individual's representation of the world» [my italics] (Wilson&Sperber 2002, p. 608).

The process of input and the derivation of cognitive effects require effort in terms of perception, memory and inference. In brief, this is the maximisation: the less processing effort required, the greater the relevance of the input (Marraffa&Meini 2005, p. 148). The subject is involved in an unconscious task of costs-benefits analysis. According to Sperber (2005), evolution would have selected cognitive mechanisms capable of foreseeing the most fruitful information with a reasonable amount of reliability. Likely, such cognitive mechanism(s) will select the easiest information to compute. In this regard, ostensive cues may serve the function of signalling the potential presence of relevant information. Moreover, the cognitive component of relevance theory is indirectly crucial for social learning context also because it entails a selective stance performed by the hearer/learner, or in other terms, an epistemic attention towards the information received both under the light of reliability of the informant source and the plausibility of the message. I will discuss this matter in more details in Chapter 6. For the moment, I would like to turn the focus on communicative aspects of relevance theory.

6.1 Recognition of communicative intention: some differences and similarities between Gricean account and communicative principle of relevance

Relevance theory provided by Sperber and Wilson tries to work out one of Grice's principle claims which are tied to our central theme: «that an essential feature of most human communication, both verbal and non-verbal, is the expression and recognition of intentions» (Wilson&Sperber 2002, p. 607). According to Grice, indeed, a speaker's utterance lies on a set of intentions: first of all the speaker has what Grice calls the «primary intention», in other words the intention to produce a certain reaction in the audience (first clause). Then, it comes the intention to make manifest to the audience the primary intention (second clause). Finally, the third speaker's intention regards the effect produced in the auditory that must be the result of the recognition by the auditory

of the speaker's intention to produce that effect (Labinaz 2012, p. 314; Grice 1957[21]) (third clause). The third clause may be simplified in the way that the hearer recognises herself to be the addressee of speaker's act.

Very briefly, according to Grice, the main feature that distinguishes a communicative act from non-communicative one is its *intentional* character. One of the Grice's philosophical effort was «to explain how the hearer infers the speaker's meaning on the basis of the evidence provided» (Wilson&Sperber 2002, p. 607), and Grice himself worried that his theory requested too sophisticated skills «to be found in a language-destitute creature» (Grice 1986, p. 85). Indeed, the communicative interaction drawn by Gricean account should *presuppose* a suite of socio-cognitive skills including i) concept of belief, ii) the capacity to infer others' goal-directed behaviours, and also iii) the capacity to sustain high levels of metarepresentation (Moore 2016b, p. 303). Such three capacities must be already possessed (and not be developed) for entertaining a Gricean communication. Subsequently, it would be impossible that preverbal infants interact communicatively in that way sustaining so complex inferences. In this regard, Dorit Bar-On (2013) even provided a non-Gricean construal of expressive communication for nonhuman animals that could be adequate to some extent for explaining communicative interchange competences among preverbal infants too. In fact, during their development, typical human linguistic expressive vehicles would come «to replace, augment, and transform the nonlinguistic expressive means» (Bar-On 2013, p. 342), such as «gaze behaviour, emotional expressions, and emotionally charged body postures and intonations that give […] direct, non-inferential knowledge of others' mental states» (Moore 2016b, p. 312).

Sperber and Wilson (2002), on their turn, hypothesise that the computation of cognitive modules would allow children's facility for Gricean communication. The main congruent aspect of Gricean theory of meaning and the notion of intention developed by relevance theory consists in the concept of *expectations* created by communicative (verbal and not-verbal) acts of the speaker. Such expectations «guide the hearer toward the speaker's meaning» (Wilson&Sperber 2002, p. 607). The use of ostensive signals in a conversational relation facilitates the communicator to provide

[21] See also Grice 1968; 1969; 1989. For a review see Labinaz 2012, and the entry "Paul Grice" by Grrandy R., Warner R. (2013) in the Stanford Encyclopedia of Philosophy.

evidence «not only for the conclusion she intends the addressee to draw, but also of the fact that she intends him to draw this conclusion. This is 'ostensive inferential' communication proper: that is, communication achieved by ostensively providing an addressee with evidence which enables him to infer the communicator's meaning» (Sperber&Wilson 2002, p. 24). Specifically, the ostensive inferential communication is formed – according to Sperber and Wilson – by two kinds of intention: «the intention to inform an audience of something» (*informative intention*), and «the intention to inform the audience of one's informative intention» (*communicative intention*) (Wilson&Sperber 2002, p. 611).

The aspect tied to an inferential process guided by mutual expectations represents also the crucial point of estrangement between the two theories. In fact, if «Grice described these expectations in terms of a Cooperative Principle and maxims of Quality (truthfulness), Quantity (informativeness), Relation (relevance), and Manner (clarity), which speakers are expected to observe [Grice 1989]» (Wilson&Sperber 2002, p. 607), Sperber and Wilson reduced the expectations only to maxim of *relevance*. Within their theory, the gricean maxim of relevance becomes an innate psychological principle which guides human cognition (Labinaz 2012, p. 337).

The cognitive principle of relevance is closely connected to communicative intentions and to ascription of others' mental state. Sperber and Wilson (1986; 1995; 2002) argued that the cognitive tendency to maximise the relevance «makes it possible to predict and manipulate the mental states of others» (Wilson&Sperber 2002, p. 611). If we knew the universal tendency to gather up the most relevant inputs, we would be able to produce a stimulus which attracts others' attention, activate proper contextual assumptions and address them to an intended conclusion. Therefore, Sperber and Wilson can claim that from the cognitive principle of relevance derives that ostensive stimuli, «designed to attract an audience's attention and focus it on the communicator's meaning», create expectations of relevance (Wilson&Sperber 2002, p. 611).

> «Given the cognitive tendency to maximize relevance, an audience will only pay attention to an input that seems relevant enough. By producing an ostensive stimulus, the communicator therefore encourages her audience to presume that it is relevant enough to be worth processing» (Wilson&Sperber 2002, p. 611).

In other words, the ostensive stimulus conveys «a *presumption* of its own optimal relevance» (Mazzarella 2016, p. 181), as the *communicative principle of relevance* claims. Such presumption of optimal relevance constitutes the main aspect of the relevance theory that has been transferred within the natural pedagogy theory by Csibra and Gergely (2006). Indeed, extending relevance theory to nonverbal communication, Gergely and Csibra posit, on the one hand, that ostensive cues deployed (also) in non-linguistic communicative context enable the agent (teacher) to communicate a «message destined to influence the targeted recipient but also the very fact that this message is being intentionally communicated to her» (Gergely&Csibra 2009, p. 149). On the other hand, the ostensive cues elicit in the addressee (learner) the «presumption of relevance», i.e. infant's disposition to interpret the transmitted message as containing «*novel and relevant knowledge*» (Csibra&Gergely 2006, p. 256).

Richard Moore (2016a, p. 43) pointed out how relevance and natural pedagogy theories «work in unison. [Learners] recognise that their teacher [is] acting with communicative intent on the basis of her addressing them with ostensive cues». In Csibra and Gergely's words:

> This aspect of pedagogy […] is analogous to the communicative principle of relevance in verbal communication (Sperber&Wilson 1986) in that it provides guidance for the learner in figuring out the knowledge content that he is supposed to acquire by the teacher's communication (Csibra&Gergely 2006, p. 256).

Going back to our tale about the baby who observes the man, we can claim that it is the very ostensive communication of the agent in front of her that determines her particular attention and the presumption that what he is going to do will be worth being attended to. Southgate, Chevalier and Csibra (2009) explicitly attempted to show whether selective imitation of a demonstrator's actions in three groups of 18-month-old infants «may be based on the same search for relevance that drives adult interpretation of ostensive communication» (Southgate et al. 2009, p. 1013). Their results «indicate that, like adults, human infants expect communication to contain relevant content, and imitate action elements that, relative to their current knowledge state or to the common ground with the demonstrator, is identified as most relevant» (Southgate et al. 2009, p. 1013).

Sperber and Wilson argue that the inferential interpretation process about communicative intentions «involves a dedicated comprehension module with its own special principles and mechanisms» (Sperber&Wilson 2002, p. 3). They sustain that the mindreading system grounds the comprehension of communicative intention, and given «the complexity of mindreading, the variety of tasks it has to perform and the particular sub-regularities they exhibit» (Wilson 2005, p. 312) – they claim that:

> «it is reasonable to assume that mindreading is not a single, relatively homogeneous system but a collection of autonomous mechanisms or sub-modules articulated together in some way» (Wilson 2005, p. 312).

For our purpose, the question is to determine whether among such sub-modules there is that one able to allow the recognition of communicative intention, and thus, «an automatic application of a relevance-based procedure to ostensive stimuli» (Sperber&Wilson 2002, p. 30) already in the early infancy. The dilemma is still very hard to settle, as Richard Breheny (2006) had already noted, offering «a minimalist account of communication, drawing on ideas from relevance theory and situation theory» (Breheny 2006, p. 74). The dilemma he lucidly raised consists in the fact that «competent language users have the conceptual abilities to […] make folk-psychological inferences about agents [speakers]», and this conflicts with one of the current developmental psychology account according to which «children below the age of four years do not possess these abilities[22]. The conflict arises because it is widely agreed in research on language development that children below the age of three years are competent language users and communicators in the basic sense» (Ibidem). As I hope it will be evident in the course of this dissertation, young children seem to be capable to entertain forms of precocious mindreading (i.e., attribution of epistemic states including false beliefs). However, our starting assumption is that the attribution of communicative intentions to others represents the *first step* for acquiring forms of cultural knowledge (including first words).

My proposal for a partial revision of natural pedagogy theory needs to cooperate with raw forms of false belief attribution identifiable with propositional attitudes in

[22] See chapters 4 and 5 for the illustration of such developmental accounts.

order to permit the presumption of omniscience bias[23]. I suggest that primary forms of mindreading are more directly connected to such kind of expectations rather than recognition of communicative intentions which can be decoded probably by other deep cognitive mechanisms, as I am going to show in the next chapter. For the moment, the intent is to find advocates of the hypothesis that, at very early stage of their development, infants hold the ability to recognise communicative intentions that trigger a cognitively specific communicative interaction.

6.2 Neo-Gricean recipe

Robert Thompson (2014) attempted to outline what kinds of psychological states are required by Gricean accounts, and concluded that Neo-Gricean accounts require a substantial understanding of communicative intentions that seems to be present at an early age (Thompson 2014, p. 171). Thompson (2014, p. 185) acknowledges that early theory of mind[24] abilities «do not bear directly on the presence of communicative intentions, [however,] many of the capacities involve an understanding of what people are intending to do by performing an action, an understanding that people's knowledge depends on what they perceive, and that people's knowledge of the world can be different from the child's own knowledge».

The Neo-Gricean theorist Richard Moore (2014; 2016b) has recently provided «a functionalist re-reading of Grice» according to which there are «minimally Gricean» communicative acts which can be performed by juvenile subjects although

i) children lack concept of belief (as it has been sketched by Donald Davidson whose point of view will be discussed in the present dissertation at chapter 5);
ii) they cannot make sophisticated inferences about others' intentions; and, thus,
iii) they are unable to entertain what Sperber (2000, p. 125) called «fourth-order metarepresentational beliefs»[25].

[23] The omniscience bias can be expressed as follows: "If I learn something, I believe that everyone knows it too".

[24] I have chosen to adopt mindreading and Theory of Mind (TOM) as equivalent terms. In footnote 1 in Chapter 4 I explain the reasons of this choice.

[25] According to Sperber's Gricean account interpretation, communicators are committed at least with fourth-order metarepresentations as in the following schema: (*First order*): S(peaker) intends that -

Producing a sign[26] with an act of address should be crucially sufficient for respecting Gricean intentional structure[27]. The advantage of Moore's proposal is that forms of Gricean communication are suitable for simple communicators which are not «capable of all varieties of informative communication», but they are limited «only to direct one another's behaviour, including their gaze or attention», lacking «any understanding that beliefs can be false», and additionally naïve communicators «can entertain and identify in others only a limited range of goal-directed behaviours» (Moore 2016b, pp. 315, 317, 324). However, I do not see a consequentiality in Moore when he posits that a production of a sign together with the act of address satisfies the Gricean intentional structure of communication, and at the same time he claims that minimal forms of Gricean communication «are possible for subjects capable of entertaining only first-order metarepresentations». In virtue of the latter claim, during a communicative exchange, from infant's perspective, we would have the following schema:

«*first order*: You [the adult agent] intend that (p)

representation: I attend and respond to your gesture» (Moore 2016b, p. 320).

This schema is akin to a blind command. On the contrary, I suggest that in order to take into account the infant understanding of referential nature of ostensive acts (i.e., all the nuances of Gricean non-natural meaning of communicate acts), we need to consider the three passages of the following schema:

1) Speaker/Agent (S) produces a *sign* with the intention to provoke:

2) Hearer/Addressee (H) to produce a particular response *r*, and

3) H recognises that S intends (2).

(Second Order): H(erear) believes (or knows, sees, etc.) that - (*Third order*): S intends that - (*Fourth order*): H believes (or knows, sees, etc.) that (*Representation*) *p* (Sperber 2000, p. 125).

[26] «Typically signs are words or gestures, but they can also include facial expressions and other body parts, and may be distributed over a combination of these things». These signs may not always be used communicatively (Moore 2016b, p. 318).

[27] Acts of address are typically marked by types of ostensive cue—such as those described for instance by Csibra (2010): eye contact, directed speech, name calling, etc.

The three passages can be explained following Thompson (2014, p. 179): S intends to produce a response *r* (*representation*) in H at least partly based on the H's recognition of S's communicative intention. In other words, as an interpreter, the young child must understand that the adult agent/speaker intends to produce a response in her. Namely, juvenile and linguistically inexpert hearers try to figure out what the agent/speaker wants to communicate in virtue of the fact that they take themselves *to be the addressees* of agent's communicative act, regardless the correct interpretation (or the mere comprehension) of the message whose content can remain epistemically vague. This the crux of Csibra and Gergely try to sustain: through the recognition of communicative intentions, infants trigger a referential expectation towards the content of the message processed subsequently by evolved cognitive biases (which are termed, as we will see in Chapter 3, *generelazability* and *universality* assumptions).

7. Communicative action (interpretation)

Following explicitly Grice, Csibra wrote (2010) that «a communicator can produce signals that are specifically designed to generate the interpretation that the communicator has a communicative intention addressed to the interpreter» (Csibra 2010, p. 144). If ostensive cues indicate a communicative intention, the question that emerges concerns what cognitive system is responsible for their correct interpretation. Since ostensive cues need to indicate that the source of information is manifesting of having an «informative intention», and whom is the target of her/his intention (i.e., who is the addressee) (Csibra 2010, p. 144), according to Csibra just «a pre-wired code system» is enough to interpret them correctly.

> «Ostensive signals need to indicate only two things: (1) that the source is making manifest of having an informative intention and (2) who is targeted by this intention (who is the addressee). Conventionalized or pre-wired code systems could do this job perfectly» (Csibra 2010, p. 144).

What does it mean *interpreting correctly* ostensive signals? «Understanding communicative actions requires that the observer knows that the actor has an intention to communicate and the meanings of the actions» (Möttönen et a. 2016, p. 230).

Definitively, we must consider two inseparable elements when we deal with the full comprehension of ostensive signals: their intentional and their referential aspect. Are infants able to catch them together? According to Csibra (2010, p. 141):

> «the presence of communicative intentions can be recognized in others' behavior before the content of these intentions is accessed or inferred. Second, […] such recognition can be achieved by decoding specialized ostensive signals. Third, […] by decoding ostensive signals, human infants are capable of recognizing communicative intentions addressed to them».

For what concerns the most universal ostensive cues, like eye-contact or deictic gestures, the scientific community agrees that early infants shift their attention following the direction of another's gaze, but there is not agreement around what such behaviour can reveal about the mind of the young child (Carey 2009, p. 175). The following question formulated by Carey is crucial: «when the child follows another's gaze or point, does the child make an agentive attribution?» (Carey 2009, p. 177). Gaze-following and pointing gestures seem to indicate evidence that agents' infant representations refer to seeking or providing information about the world (Carey 2009, p. 177). As we will see better in Chapter 2, it is reasonable to suppose, following an amount of experimental data (Cecchini et al. 2011), that from the birth we may glimpse referential expectations in infant behavioural responses which mature during the first months of life. Four-five month-old infants seem to be able to interpret ostensive signals «obligatorily» as directed to them (Parise&Csibra 2013, p. 1). Such experimental data, combined with Lloyd-Fox and colleagues' study (2015), bring to the plausible conclusion that:

> «the brain of young infants produces a *quick obligatory response to the presence of any ostensive signal*, but would invest enhanced processing of the communicative acts of the source of these signals only if the nature of another potentially ostensive signal does not conflict with the interpretation that they are the ones who are being addressed by the communicator» [my italics] (Lloyd-Fox et al. 2015, p. 46).

7.1 Second-person view: some reflections

Recently, Tauzin and Gergely (2018) showed how 13-month-old infants can properly recognise a significant communicative relation involving the transmission of information from a knowledgeable agent to a naïve one. In their experiment, infants inferred that «the turn-taking exchange of variable tone sequences between unfamiliar agents [was] indicative of communicative transfer of goal-relevant information», but infants did not manage to do such inference when «no goal-relevant contextual change was observed that would motivate its communicative transmission», or when «the agents exchanged fully predictable identical signal sequences» (Tauzin&Gergely 2018, p. 1). This experiment indicates the infant high sensitivity for communicative (non-verbal) signals also between third-party agents: a sensitivity which induces them to interpret particular cues as forms of information exchange. Anyway, communicative actions imply a second-person view, namely the direct infant's involvement into the action. It appears quite evident the importance of the context to interpret the kinds of action. Bosco and colleagues (2004) showed that different communicative meanings can be assigned to the same expressive acts. Their experiments conducted on three groups of children aged 3–7 years reveal that different contexts «play different roles in the reconstruction of the communicative intentions» (Bosco et al. 2004, p. 467).

In the second-person perspective, agent performs two kinds of action: *i)* action aimed to create an ostensive-communicative context, and *ii)* action that specifically deals with the referent. *These two actions are independent*. They may be bound (causally and temporally) only for the fact that agents have the goal of teaching something about objects. This link is to be caught by the infants to trigger the *pedagogical stance*. Although referential communication does not belong to humans only, and it is spread in non-human animals evolutionally distant from our specie (Topál et al. 2008), human referential communication is unique because the agent can specify the referent separately from the content of the message. Therefore, we may say reasonably that the *triggering-moment* of natural pedagogy is marked by two developmental factors:

a) the interpretative capacity of ostensive cues, and

b) referential expectations about object (two factors which seem to belong also to other animals)[28].

In conclusion, four claims are central within natural pedagogy framework:

(1) human infants are endowed with the capacity to recognise *ostensive* signals (including eye contact, infant-directed speech, and so on) as indicating that a teacher/speaker is acting with communicative intention (Csibra 2010).

(2) On the basis of such recognition, infants feel to be the addressees of the ostensive cues, and so they look for recovering the content of the message.

(3) In virtue of the presence of ostensive cues, the addressees treat speaker's information as generalising claims about the object kind to which teacher is referring.

(4) Additionally, young children expect that what they learn is *universally* known, or in other terms, quoting directly Gergely (2007, p. 179), ostensive referential cues trigger also «*the implicit expectation* by the infant that the manifested information will contain *publicly shared universal cultural knowledge available to all others* (and not only to the demonstrator who is the communicative source of the information)» [my italics].

In this sense, Gergely can claim that natural pedagogy is a «relevance-guided social communicative learning device» (Gergely 2007, p. 173). However, before facing natural pedagogy theory and tackling in particular the point (4), it is necessary to

[28] As Gergely recognises too, ravens, dogs, scrub-jays and goats seem to have the capacity to infer referential information from other conspecifics (and humans) (Bugnyar et al 2004; Emery&Clayton 2001; Kaminski et al. 2005; Lakatos et al. 2009; Kupán 2013). Call and Tomasello (2008) explain how chimpanzee «follow gaze with referential expectations to find an object where the other is looking, and they also look behind barriers to find the object the other is gazing at» (Gergely 2011, pp. 89-90). In particular, according to Call and Tomasello (2008, p. 187): «there is solid evidence from several different experimental paradigms that chimpanzees understand the goals and intentions of others, as well as the perception and knowledge of others. Nevertheless, despite several seemingly valid attempts, there is currently no evidence that chimpanzees understand false beliefs». The last statement has been overtaken by Buttelmann et al.'s (2017) recent findings about false belief in great apes (see in this regard Chapter 4).

investigate which potential cognitive mechanism allows the recognition of communicative intention. In this way, we could provide a solution a solution to the first question sketched in the introduction: *does natural pedagogy theory claim that a communicative intention is not part of theory of mind*?

CHAPTER 2.

The infant sensitivity to ostensive communication

The present chapter has two aims. The first one is to show the great sensitivity held by infants to respond to ostensive cues. Particular attention will be given to eye-contact that represents the primary and maybe the most important ostensive signal in healthy ontogenetic development. The question is whether mindreading system is necessarily involved in such sensitivity attested in very early infants and even in newborns. The second aim of the present chapter is trying to investigate the onset of natural pedagogy system, that should coincide with the moment in which infants are able to catch the referential nature of ostensive signals. In the previous chapter, I have posited the later TOM commitment in the understanding of goal-oriented action. In the present chapter, I advance the parsimonious account about the comprehension of ostensive signalling that does not contradict the subsequent TOM commitment. In fact, the neural network involved predisposes the maturation of TOM skills in interactive communication (visual and verbal), as it is clearly attested in adulthood.

1. Communicative faces

Human beings are sensitive from birth to the presence of eyes (Csibra 2010, p. 145). Batki and colleagues (2000) showed that neonates prefer looking at faces with eyes open (Batki et al. 2000, p. 223), suggesting the presence of an innate mechanism dedicated to perceiving and detecting the direction in which eyes are looking. The researchers tested 105 neonates who were presented with two photographs separately: one depicted a female adult face with open eyes, and the other one depicted the same face with eyes closed. Neonates spent more time looking at «the photograph with the eyes open than at the photograph with the eyes closed» (Batki et al. 2000, p. 223). As Csibra (2010) noticed, if such experimental results may be explained by the presence of a neural mechanism dedicated (see below) to eye-contact, the interpretation provided, however, does not demonstrate *infants' preference* for eye contact.

More robust finding in this research pathway is, on the contrast, Farroni and colleagues' (Farroni et al. 2002) study in which newborns looked longer at faces that

engaged them in mutual gaze when compared to averted gaze. Farroni, Csibra, Simion and Johnson (2002) found that when neonates have the possibility to choose between photographs of faces looking directly at them or looking in another direction, «3-day-old newborns prefer to look at the face that appears to make eye contact with them» (Csibra 2010, p. 145). In their experiment paired photographic face stimuli were presented to seventeen newborns (7 males and 10 females). «One of the pair had direct gaze, whereas in the other face the eyes were averted randomly to the right or left» (Farroni et al. 2002, p. 9602). The results showed that neonates looked longer (mean 106.8s) at the direct gaze than at the averted gaze (mean 63.7s) (Farroni et al. 2002, p. 9602). According to the authors, the preferential attention towards direct-gaze faces provides a clear evidence that «human newborns are born prepared to *detect socially relevant information* [my italics]» (Farroni et al. 2002, p. 9602).

Their hypothesis is consistent with following findings like those conducted by Farroni et al. (2003; 2004; 2006; 2007) who tested newborns and 4-month-old infants, and found that newborns are sensitive not simply to the presence of eyes but also to the position of the pupils-irises within the eye, and more specifically to eyes with the pupils lined up centrally in a face observed from an upright position (Csibra 2010, p. 145). Therefore, in newborns the preference for faces (with direct gaze) occurs *only* «within the context of an upright face and a straight head» (Farroni et al. 2006, p. 298). On the basis of such experimental literature, face and gaze perception are strictly bound together from birth, and eye contact plays important communicative functions (Gliga&Csibra 2007, p. 323). According to Csibra:

> «canonical orientation of a face provides some extra advantage that makes it worth being preferred. Preparedness to be a recipient of communicative acts could be the extra factor that explains this aspect of early sensitivity, because, when it comes to faces, *only an upright face looking at the baby would be considered as an eye-contact stimulus*, i.e. an *ostensive signal* [my italics]» (Csibra 2010, p. 146).

1.1 Newborns prefer communicative faces

Within the context of the upright faces observed, infants and newborns seem to prefer (and they are attracted by) some faces in particular: their mother's face and new

faces (Pascalis et al. 1995; Scott&Nelson 2004[1]). However, several studies (Bertin&Striano 2006; Nagy 2008) demonstrated that both neonates and infants of few months of age decreased the looking-time spent observing a face that became frozen (still-face situation), and this phenomenon is true even for their mother's face, as Yato and colleagues showed in their study (Yato et al. 2008).

What is the main element that can really determine infant's face-gaze preference? Italian researchers of the University of Rome "La Sapienza" tried to answer this question hypothesising that the neonate looking preference at faces is guided by an «expectation of communication» (Cecchini et al. 2011, p. 425). It is an intriguing hypothesis that has important implications for understanding social competencies, because if newborns are really able to build «a dynamic social representation of the interaction with them», they could «use the representation of the interaction with the known faces in order to orient their behaviour during successive interactions» (Cecchini et al. 2011, p. 425).

Cecchini and colleagues (2011) tested 16 newborns which were divided into two groups. A motionless face (still-face) was shown for 8 minutes to Group 1, whereas Group 2 was presented a communicative face, which brought to the neonates a series of tactile, visual and vocal signal exchanges (for 8 minutes). The study aimed to measure the reactions (expressed by looking time, vocalizations, mouth movements) of infants during a preference task between a "known still- (or communicative) face" (*Known Face*) and a "new still- (or communicative) face" (*New Face*). The newborns preferred looking at the new face only when the known face was in an immobility condition (still-face). *This finding shows that newborns actively avoid looking at immobile faces.*

[1] Scott and Nelson (2004; but see also Carver et al. 2003) highlight that «as with adults, both infants and children show apparent neural dissociation between face and object processing» (Scott&Nelson 2004, p. 32). As fair summarised by Cecchini et al. (2011, p. 425) newborns have preference for mother's face maybe «because it gives positive reinforcement»; however, according to Scott and Nelson (2004) and Pascalis et al. (1995), once newborns got used to the mother's face, if we provide them a successive preference task they preferred the new face» (Cecchini et al. 2011, p. 425). Baron-Cohen (2005) highlights sex differences in several aspects of developmental behaviour due to different degree and balance of two brain activities ("systemizing" and "empathizing"), and citing the Connellan and colleagues' study (2001) he claims that there is a difference from birth between females who look longer at faces, and males who spend more time to look at inanimate objects (Baron-Cohen 2005).

Furthermore, the results indicate that when the known face was previously communicative, the newborns lost the preference for the new face (Cecchini et al. 2011, p. 431). The researchers hypothesised that the significant preference of the new face is due «to the avoidance of previously immobile and non-communicative faces» (p. 431), rather than to the habituation to a face already observed (known face). This fact makes us assume that newborns may build more complex representations of communicative faces than perceptive ones, developing communicative expectations which could guide neonates' behaviour in further interactions (Cecchini et al. 2011, p. 431).



Fig. 1 *Preference task between New Face and Known Face (Cecchini et al. 2011, p. 431)*

A subsequent research conducted by Lai and co-operators (2015) used ERP (Event Related Potential)[2] technique and temporal connectivity analysis to test whether infants before the third month of age were able to recognise a face previously seen. Previous findings, as, for instance, the influential study of Johnson (2005) hypothesised a subcortical and cortical rout for face processing. The former should involve the superior colliculus, pulvinar and amygdala for face detection, and the latter fusiform gyrus and inferior occipital gyrus for face identification (Johnson 2005, p. 767). Johnson suggested that subcortical route already functions in newborns and «not only

[2] ERP allows to measure processing information between a stimulus inducted and a response, and it is one of the most reliable method adopted in cognitive neuroscience to study neuronal correlates of perceptual and cognitive activity (see e.g., Liao et al. 2015). This technique is better than fMRI in developmental studies because it is easier to use with infants.

detects the presence of faces, and orients the newborn towards them, but might also activate relevant cortical regions such as the lateral occipital, fusiform and orbitofrontal cortices» (Johnson 2005, p. 770). Lai and coworkers (2015) confirmed for the first time such neural correlations from birth combining ERP data and connectivity analysis relatively to those brain areas supposed to play a crucial role in adult face-processing. experimenters presented a face for 1 minute (*Target*) to 23 newborns. Then fifty trials of *Target*, fifty trials of *Unknown faces* and fifty trials of a *neutral stimulus* were performed, and each trial lasted 2 seconds (Lai et al. 2015, p. 96). ERP analysis highlighted «a difference amplitude in response to *Target* vs. *Unknown* on left occipitotemporal montage»; furthermore, connectivity analysis showed «higher implication in fusiform gyrus with known face» (*Ibidem*). These results together induce to claim that newborns may «discriminate a familiar face from a stranger face since birth» (*Ibidem*).

1.2 What mechanism is involved for early gaze detection?

Once such early capacity has been attested, it is important to determine the onset of gaze detection that represents the necessary further step for the establishment of a communicative relationship. Two candidate mechanisms may underlie newborns preference for direct eye gaze. One is referred to Baron-Cohen's proposal about the multi-componential structure of mindreading system (Baron-Cohen 1994; 2005). According to his original account, mindreading faculty develops little by little during the first 2-4 years of age through three fundamental stages in which several mechanisms start to interact. According to Baron-Cohen's revisited model for the ontogeny of mind system theory, during the first 9 months of age human infants trigger three innate distinct mechanisms, EDD (Eye Direction Detector), ID (Intentionality Detector), and TED (The Emotion Detector) (Baron-Cohen 2005). The model is termed by the author «Empathizing System», and it is more adequate than the previous one (proposed in 1994) to describe the computational processes of «an observer impelled towards action», because it involves also «information about affective states, available to the infant perceptual system» (Baron-Cohen 2005, p. 8). EED, ID and TED build the dyadic representation of simple mental states (Baron-Cohen 2005).

Keeping in mind the beginning example of the man who entered the room, we may say that EED makes the infant able to represent man's eye stimuli in the propositional form "he is looking at me" and then "he is looking at something else there". The amodal module ID serves the function to represent man's agency as goal-directed in the propositional form like "he wants to eat", and finally TED represent the affective state in the propositional form "he is happy to eat the soup"[3].

The alternative candidate mechanism underlying the preference for direct eye gaze is provided by Johnson and Morton (1991a; 1991b), who hypothesised «that subcortical circuits supported a primitive representation of high-contrast elements relating to the location of the eyes and mouth» (Farroni et al. 2002, p. 9604). According to Johnson and Morton (1991a), information about face structure is hard-wired in newborns' mind. They defined such structural information «CONSPEC», that would guide the preference for face. The primitive system CONSPEC would be located in subcortical structures (superior colliculus, pulvinar, amygdala), whereas the later system termed CONLERN, deputed to acquire and retain «specific information about the visual characteristics of conspecifics», would be mediated by cortical functions (Johnson&Morton 1991a, pp. 172, 175).

Farroni and colleagues (2002) support Johnson and Morton's account considering the fact that around the fourth month of age the presence of direct gaze facilitates neural processes associated with the earliest face encoding ability[4]. Actually, in a prior study, Teresa Farroni and colleagues (2000) challenged the possible presence of a specific module dedicated to eye gaze detection[5], and in further experiments Farroni, Mansfield, Lai and Johnson (2003) criticised directly the EED hypothesis,

[3] The following developmental step is the onset of Shared Attention Mechanism (SAM), and it should come on line at the end of the first birthday until 14 months of age. SAM combines ID and EDD and builds a triadic representation expressible in the form: "Man (and also my mother) see that I see the soup in the plate". SAM would be the mechanism that allows joint attention (Baron-Cohen 2005, p. 4). The third developmental step is the maturation between the second and fourth year of age of theory of mind module (TOMM). It allows for the pretend play (as described accuracy for the first time by Leslie (1987)), the ascription to others of false-belief, and the understanding of the relationships between mental states (as proposed by Wellmann (1990)).

[4] For further sustain to CONSPEC-COLERN account see Johnson 2005, and Gliga&Csibra 2007.

[5] See also Johnson 1994.

suggesting, in contrast, the activity of a domain-general process rather than a domain-specific module as predicted by Baron-Cohen.

Farroni and co-operators tested 4-month-old infants and found that the attentive mechanisms that make infants able to follow eye gaze shift are engaged *only by a previous mutual gaze* with an upright face (Farroni et al. 2003, p. 209). Therefore, the authors noticed that if «a period of mutual gaze with an upright face facilitates the direction of attention» (Farroni et al. 2003, p. 210), EED hypothesis fails its predictions under two respects: *i*) infants are not automatically sensitive to the direction of eye gaze; *ii*) there is no clear evidence that 4-month-olds at least build a dyadic representation in the form described above.

However, the development implies an increase of specificity in detection of eye-gaze-direction ability, as attested in childhood and evident in adulthood. Within the ontogenetic perspective, this fact may be the result, as suggested by Johnson and Farroni (2003; see also Farroni et al. 2003), of interactions of specialised brain areas. In this regard, for what concerns 4-month-old infants, tests show that the direction of the gaze modulates face recognition also in adults (Farroni et al. 2007). Under what aspects are infants similar to adults in ostensive (face) stimuli recognition?

1.2.1 Adults and infants: similarity or not similarity?

Grossmann and colleagues, through gamma band oscillatory brain activity, examined the neural basis of 4-month-old infants' perception of eye gaze direction. Photographic images of upright and inverted faces directing infant's gaze towards them or to the side were presented.

> «Direct gaze compared to averted gaze in upright faces elicited increased early evoked gamma activity at occipital [and right prefrontal channels], suggesting that eye contact detection might recruit very similar cortical regions as in adults» (Grossmann et al. 2008a, p. 282).

Specifically, in another study, Grossmann and colleagues (2008b) investigated whether the involvement of cortical areas in adults' perception of facial communication signals was already functionally active in early human development. Their results

intend to show that facial communication signals activate areas in the infant temporal and prefrontal cortex corresponding to the brain regions implicated in the same process in adults. Therefore, they conclude, there is «an early specialisation of cortical network involved in the perception of facial communication cues, *which is essential for infants' interaction with, and learning from, others* [my italics]» (Grossmann et al. 2008b, p. 2803; Grossmann et al. 2010). However, we must be cautious, because there is a risk to make a perspective's mistake confusing the overlapping phenomena due to the neurobiological and cognitive development, and the real underlying neural basis. Neuroscience teaches us that there is a network of brain structure in human social interaction and communication. Such network, universally known as "the social brain", is constituted by subcortical and cortical regions, as superior temporal gyrus, fusiform gyrus, cingulate gyrus, ventral and medial prefrontal cortex, and amygdala. All these regions are specialised and involved to process several and distinct social information.

For our purposes, the difficulty is to discriminate the primary areas and mechanism really involved in processing ostensive signals during early infancy. The mistake to avoid is attributing the process of detecting eye contact, for example, to the brain regions relevant to theory of mind (TOM) only in virtue of the fact, as Senju and Johnson (2009) noticed, that the subcortical and cortical regions involved in TOM overlap with areas dedicated to eye contact[6]. However, the involvement of mentalization may not be useful for understanding Senju and Johnson's (2009, p. 127) term "*eye contact effect*" (EFC), that indicates the phenomenon for which «perceived eye contact modulates the concurrent and/or immediately following cognitive processing and/or behavioral response» (Senju&Johnson 2009, p. 127).

In order to explain better ECF, Senju and Johnson propose the alternative model called "the fast-track modulator", that develops Morton and Johnson's proposal. Senju and Johnson's model predicts that «eye contact effect is mediated by the subcortical face detection pathway» (including superior colliculus, pulvinar, and amygdala) (Senju&Johnson 2009, pp. 129-130). Such neural route operates rapidly and modulates face processing, and it is the responsible, as proposed also by Johnson (2005), of face preferences in newborns. In other words, the subcortical pathway would underlie eye

[6] This kind of attribution has been made by Kampe and colleagues (2003), but in the case of adults and not of newborns or infants, as we will see in § 2.3.1.

contact effect and would precede the full maturation of cortical processing of gaze direction.

In my opinion, Senju and Johnson's proposal provides the best model able, on one hand, to not backdate mentalistic skills to birth for the preference of and the sensitivity to eye detection, and on the other hand, the model allows to explain the subsequent overlapping of TOM, given «the widespread connections between the subcortical route and cortical structures» (Senju&Johnson 2009, p. 132).

1.3 *Tactile communication in newborns*

The preference for communicative cues in newborns is evident also if we change the communicative channel by which the signal is perceived. With the background assumption that infant crying behaviour serves an important role in the communicative process between caregivers and newborns (Michelsson 2001; Barr et al. 2000), Cecchini, Lai and Langher (2007) investigated the potential relation between three different communicative contexts characterised by the absent, continuous or discontinuous tactile communication, and the newborns' crying behaviours (coded through the features of latency, duration and dysphony[7]).

They tested thirty newborns randomly divided into three groups. Each group of ten newborns paired by gender was presented with two visual stimuli, a still-face (social-stimulus) consisting of «a tri-dimensional plastic woman's face (22cm×15cm) with brown hair», and a square (non-social stimulus) consisting of «a grey square (22cm×15cm) with a black frame». Each presentation was for a duration of twelve minutes with an interval of three minutes. Only in group 2 and group 3 the visual stimuli were preceded by a tactile communication. An experimenter, who stood behind the cradle that the neonates could not see, performed a tactile communication for eight minutes presenting her finger to each newborn who held it. The experimenter's slight movements with her finger were *continuous* with group 2, and *discontinuous* with group 3 (Cecchini et al. 2007, pp. 657-658).

[7] Dysphony (or turbulence) is intended here as a portion of Expiratory Cry Tract (ECT) «characterised on the spectrogram by non-equidistant and often not well-defined harmonics […] and acoustically by a raucous, rough or harsh sound» (Cecchini et al. 2007, p. 658).

The authors noticed that the presence of tactile communication affected the amount and the quality of crying behaviour. Indeed, during the tactile experience in groups 2 and 3 «the percentage of crying was near to zero» (Ibid., p. 663), suggesting that such experience was pleasant. On the contrary, during the visual stimuli presentation (still-face and grey square) every group cried, but with important differences. The crying manifestations were less frequent in Group 2, and no significant crying episodes occurred in Group 3 during the whole experiment. Starting from these results, the authors concluded that «the presence of tactile communication […,] compared to its absence (group 1), predisposed newborns to cry later, less, and with less dysphony […] during the stimuli presentation» (Ibidem).

Now it is worthwhile making two important considerations. The first one regards the absence of significant differences between the crying reactions in front of the social and non-social kind of visual stimulus. Newborns cried both in front of a still-face and grey square; no difference was found by the experimenters. Infants considered a still face as "negative" as a neutral object.

The second consideration is about the different degree and intensity of crying between group 2 and 3. Why did the newborns belonging to group 2, who received continuous tactile communication, cry more than those who received the discontinuous tactile experience (group 3)? This event is explained by the authors in terms of communicative expectations exchanges. Indeed, for the fact that newborns in group 2 were used to a steady pleasant tactile communication, their crying during visual stimuli presentations did mean the need «to recall the steady exchange of signal» (Cecchini et al. 2007, p. 663). Instead, newborns in group 3 were used to staggered tactile exchange, thus, they did not cry (so much) «because they expected that in some periods no communicative exchanges occurred» (Ibidem).

Evidence seems to indicate the innate attitude held by infants from birth to reach communicative signals, maybe independently from the kind of communicative source, and the communicative channel by which the communicative cue is perceived, as suggested by Parise and Csibra (2013) for older infants (5-month-old). For a full-blown ostensive communication it is necessary that infants understand not only the referent targeted by the gaze direction but, first, that the very ostensive signals are addressed to

them. Therefore, we have to analyse two crucial aspects of infant ostensive communication:

1) *when* and *whether* they understand that the very manifestation of a communicative (speech) act refers to something, i.e. has primly a *communicative function* that can be, thus, «separable from its lexical content», just like those utterances, for example that have the form of speech and that «can be interpreted as communicative even when the content cannot be comprehended» (Vouloumanos et al. 2014, p. 872; see also Martin et al. 2012; Vouloumanos et al. 2012).
2) *when* and *whether* infants understand to be the addressee of the communicative act.

The point 1) is important also for investigations about language acquisition, but it is not the topic here. Rather, it is more significant to report those data attesting the infants' understanding of communicative function of speech. Vouloumanos and colleagues (2014) tested «whether 6-month-old infants recognize that a Communicator can inform a Recipient about a target object by using speech (a novel speech token, 'koba') but not non-speech (a coughing sound, 'xhm-xhm-xhm')» (Vouloumanos et al. 2014, p. 873). In other words, in their experiment, an infant observes communicative exchanges between a *Communicator* and a *Recipient*, and in order to understand «the Recipient's response to the speech, the infant must infer that the speech can convey information to the Recipient, even though the novel word has no established meaning for the infant» (Ibidem). Here, infants are mere observers and are not involved in the supposed exchange of information. This permits to analyse the objective understanding of the communicative function of a linguistic act regardless of the content. In the experiment abovementioned, that used looking time paradigm, 6-month-olds showed to have expectation regarding the form of speech («a novel speech token not yet associated with any established meaning») to allow the Communicator to convey relevant information to the Recipient (Ibid., p. 878). In other words, the results indicate that, by 6 months, «infants understand that the form of speech, independent of any specific lexical content, can communicate information about an object» (Ibid., p. 877).

The point 2) is the matter of the following paragraphs.

2. *Referential expectations*

2.1. *Eye contact influences the processing of objects in 4-month-old infants*

Infants are guided, roughly said, towards the object-target by the adult's gaze. Following eye gaze and shifting attention to the direction of an adult's eye gaze represent fundamental abilities for the maturation of joint attention (Moore&Dunham 1995), which are crucial for imitative learning in infants, as it is universally acknowledged (Brooks&Meltzoff 2005; Striano&Stahl, 2005; Striano et al. 2006a; Csibra and Gergely 2006), and determining also for other cognitive processes like language development (Tomasello 1995; Brook&Striano 2005).

In former times, it has been sustained (Carpenter et al. 1998) that only after 9 months of age infants are able to engage in joint attention. More recent studies challenged this view. In this regard, Hoehl and colleagues (2008) tested seventeen 4-month-old infants showing them «static photographs of faces with eye gaze averted to the left or right side, with one object presented near the face» (Hoehl et al. 2008, p. 11). The results of their experiment suggest that infants process object-directed eye gaze faster than non-object-directed gaze. This is consistent with the previous Reid and Striano's study (2005) according to which «adult eye gaze biases infant attention» (Reid&Striano 2005, p. 1765). In a few words, 4-month-olds' different processing of objects depends on whether the objects were cued by the direction of an adult's eye gaze. Furthermore, Hoehl and colleagues claim that infants are able to «encode socially cued information» with more accuracy than not-socially cued information (Hoehl et al. 2008, p. 15). Therefore, information transmitted through social interactions is better processed by such young infants than information provided by non-social sources.

It would be interesting to test, in further researches with very early infants, whether providing non-social informative sources yet equally communicative rendered through technical device determines the same results[8]. Anyway, Vouloumanos et al. (2014) and Hoehl et al.'s (2008) findings just mentioned induce to think that infants see and look for other people around them not only as a source of nutritive and emotional

[8] See Deligianni et al.'s (2011) study on 8 month-old infants illustrated at the end of the present chapter.

care, but also as a source of information about the surrounding environment, and as guides who allow them to learn about natural and artefact objects.

We are assuming, in this way, a sort of *innate epistemic attitude* that projects infants towards the world like spontaneous searchers, or hunters of information and knowledge potentially obtainable by any kind of communicative sources able to catch first and shift then their attention towards a referent.

2.2. *Eye contact influences the processing of objects in 5-month-old infants*

The experiments reported in the previous paragraph used the ERP technique (Event Related Potential). Event related potentials are the most appropriate markers for measuring cognitive processes when tasks request passive observation by infants rather than evident behavioural responses. ERPs are the most appropriate techniques for investigating cognitive processes in infants because they result less invasive than fMRI. ERP allows to measure processing information between a stimulus inducted and a response, and it is one of the most reliable methods adopted in cognitive neuroscience to study neuronal correlates of perceptual and cognitive activity. ERP is constituted by underlying components which are related to waveform mark that has a series of positive and negative voltage deflection. Parise and colleagues (2008) focused on measuring the so called "middle latency negative central component (Nc)"[9], that «has demonstrated sensitivity in experiments involving eye gaze or joint attention» (Parise et al. 2008, p 143). Parise with his group of researchers investigated the neural effect of joint attention on object processing during live interactions between adults and 5-month-old infants, who were divided into two blocks. Each block included a pretest phase in which a female experimenter uttered short phrases like: "Oh nice!", "So many colours", with friendly facial expressions, smiles, and a positive tone of voice, while she was sitting in front of the infant and turned her head alternately from the baby to a screen, on which three different objects were presented. The only difference between the two blocks was in the kind of familiarisation, i.e. the nature of eye contact that the experimenter engaged with the infant during the pretest phase (Parise et al. 2008, p. 147): in "joint

[9] It appears approximately 300-700 ms after stimulus onset and it is most prominent at fronto-central electrodes (Webb et al. 2005).

attention condition" there was mutual eye contact between the experimenter and the infant, while in "no-joint attention condition" the experimenter did not look directly at infant's eyes.

The interpretation provided by Parise and colleagues (2008) to ERPs recorded are consistent with the abovementioned findings on 4-month-olds, highlighting the strong influence on infant object processing of a period of mutual eye contact (Parise et al. 2008, p. 148). It is worthwhile underlining that their data interpretation is also comparable with Senju, Johnson and Csibra's (2006) study and Senju, Csibra & Johnson's (2008) experiments on joint attention in 9-month-old infants. Both findings used ERP measures, but the former is intriguing because the authors confronted 9-month-olds with adults, reaching the conclusion that the different subjects «encode referential information of gaze in a similar way» (Senju et al. 2006, p. 220). Now, Parise and colleagues noticed that there is a great similarity in Nc ERP component between 5 and 9-month-olds (as it is evident also in Striano et al. 2006b)[10], suggesting «the possibility that the neural systems subserving the extraction of information during social interactions are the same at both ages» (Parise et al. 2008, p. 148).

2.3."There is something for me!"

It seems to be clearer, at this point, that in infants there is a rudimentary comprehension of the referential nature of gaze shift, that is facilitated by previous mutual eye contact. Senju and Csibra (2008) found in 6-month-old infants that such facilitation is more evident if the act of following adult's gaze shift is preceded by more ostensive signals joint together, as eye contact plus infant-direct speech (IDS), termed also "motherese"[11].

The evidence provided by Csibra and Volein's (2008) finding suggests that infants of only 8 months of age *expect a referent object for a gaze shift*. They tested

[10] Striano and colleagues (2006b) found in 9-month-old infants that the Nc increased in amplitude during the processing of objects during joint attention interaction with adults compared to non-joint attention condition.

[11] Adults alter their speech by specific vocalisations, heightened pitch, exaggerated pitch excursions, short utterance length limited to shared experience, increased of repetition, and other specific features in order «to facilitate infants' meaningful processing of the speech stream» (for a review see Futó 2012, p. 8).

infants who expected an object to be at a location indicated by someone's gaze even if the object was hidden by an occluder. This has an important implication, because, as suggested by Csibra and Volein (2008; see also Csibra 2010), it could mean that infants can infer that the mutual eye contact and the following gaze shift imply a message transmitted by the communicator in front of them, and such message is the referent.

Going back to our example, we should imagine that the young child - observing the man who is now watching something in the kitchen after long mutual eye contact and other communicative exchanges – thinks something as: "Oh, there is something for me there!". Taking the plate and starting to taste the soup, the man will satisfy his/her referential expectation.

Given the experimental results discussed above, and the intriguing relationship between 5 and 9-month-old infants for what concerns joint attention, the question that arises is whether 5-month-olds at least are able to understand that targeting an object through gaze shift, and also other accompanying ostensive stimuli, implies a message *for* themselves, i.e. whether 5-month-olds, or younger infants, feel being addressed by the message.

2.3.1 Responses to multimodal ostensive signals

Three experimental studies (Grossman et al. 2010; Parise&Csibra 2013; Lloyd-Fox et al. 2015) may help answer this question. The main element which joins all these findings is the researchers' attempt to investigate the neural basis of detecting *multimodal ostensive signals* in early development, focusing in particular on eye contact and IDS (Parise&Csibra 2013; Lloyd-Fox et al. 2015), and eye contact associated to detecting process of infant's own name, a spoken stimulus always belonging to IDS (Grossmann et al. 2010). According to the authors, the combination of stimuli from different modalities and perceptual channels could help to analyse more deeply the nature of cognitive response that ostensive signals determine in early infants.

There are several ostensive signals that can be combined. Indeed, the infants' preference of motherese is observable in the domain of actions too, that are termed for this reason *infant direct-action*, or "motionese" (Brand et al. 2002) often accompanied by motherese speech. In this regard, for instance, Brand and Shellcross (2007), in a

preferential looking paradigm tested 6- to 8-month olds and 11- to 13-month-olds who «were shown to have a systematic preference for infant-direct-action over adult-direct action. […] [The] results indicate[d] that infant direct modifications involving hands, arms, bodies, and/or objects can sufficiently engage infants' attention even without the presence of corresponding communicative facial expression and eye-gaze» (Futó 2010, p. 12).

Around the fifth month of age, and even before, in addition to following gaze shift, infants start to learn other ostensive signals such as their own name (Mandel et al. 1995; Parise et al. 2010). In order to investigate «how 5-month-old infants process their own names, […] and how one's name enhances infants' attention to objects» (Parise et al. 2010), Parise, Friederici and Striano used ERP methodology to compare neural response in infants when they hear their own name or a stranger's name whereas they look at novel objects. The results of this finding indicate that «hearing her own name prepares the infant to receive new relevant information» (Parise et al. 2010). Grossmann, Parise and Friederici (2010) examined, through NIRS (Near-Infrared Spectroscopy)[12] technique, 5-month-old infants engaged in watching human faces while listening to voices which called them with their own name or another name. More precisely, during the experimental setting, pictures of smiling human faces depicted with mutual and averted gaze were presented while the infants sat on the parent's lap and heard their own name, or a stranger's name, uttered by a female voce (in infant-direct-speech (IDS) modality) transmitted by audio files.

The researchers found that «infants recruit adjacent but non-overlapping regions in the left dorsal prefrontal cortex when they process eye contact and own name» (Grossmann et al. 2010, p. 1). As it is shown in the picture below, there are two adjacent

[12] «The NIRS method relies on the optical determination of changes in hemoglobin concentrations [and oxygenation] in cerebral cortex which result from increased regional cerebral blood flow» (Grossman et -al. 2010, p. 2). NIRS data can be compared with fMRI data even if the resolution is not so good but it represents a compromise given the complexity to test infants lying on a tube. Indeed, thanks to NIRS, infants wear a lightweight hat that shines a light across the scalp, and measures the amount of light absorbed. The absorption level variations will depend on the volume of hemoglobin in that area (Redcay&Saxe 2013, p. 230). It is considered also «a viable procedure for assessing the relation between object processing and brain function in human infants» (Wilcox et al. 2005). For more details about NIRS methodology applied in developmental studies see Lloyd-Fox et al. (2010).

left dorsal prefrontal regions significantly sensitive to two ostensive signals compared: that indicated by bottom 6 responds better to eye contact, whilst the other one (bottom 9) shows more sensitivity to infant's own name (Grossmann et al. 2010, p. 3). Grossman and colleagues' experiment is important for two reasons:

i) the results provide further support to Senju and Johnson's model (Senju&Johnson 2009), and Johnson's (2005) hypothesis about the presence of a rapid subcortical pathway (involving the superior colliculus and the pulvinar) that conveys information to the prefrontal areas. Indeed, the infant response to eye contact recorded in prefrontal cortex was significantly faster than the response to the IDS cue (Grossmann et al. 2010, p. 4).

ii) There is opposite lateralisation of the prefrontal cortex areas involved in adults for similar tasks. Infants lack the right lateralisation present in adult processing of the same ostensive cues, as the Kampe and colleagues' (2003) experiments showed.

In particular, their experiment (Kampe et al. 2003) attempted to provide neural basis to Sperber and Wilson's (1995) relevance theory, and specifically, Kampe, Chris and Uta Frith (2003) wanted to test whether adults involved with ostensive acts would activate those brain regions implicated in mentalising. They found that the social brain areas relatively to TOM competencies were recruited both in detecting eye contact, and when subjects were called by their own names. According to their study, the areas involved in eye contact are «the paracingulate cortex, the temporal poles, and the temporoparietal junction», whereas the other ostensive cue (calling the person's own name as opposed to a different name) activate significantly «the right paracingulate cortex and the right and left temporal pole, the medial surface of the superior frontal gyrus, and the inferior frontal gyrus/insula on both sides» (Kampe et al. 2003, p. 5260). The conclusions of Kampe and colleagues (2003) were sharp: *the comprehension of the two simplest and most common ostensive cues is due to neural circuits underlying mindreading system*. This would confirm perfectly Sperber and Wilson's predictions. Therefore, relatively to adults at least, without mentalizing, the recipients of the communicative act would not recognise the nature of ostensive signals as referring to

themselves (Kampe et al. 2003, p. 5262). Another study (Grossmann&Johnson 2010) has examined triadic attention in infancy with NIRS testing 5-month-olds. The researchers found that the left dorsal medial prefrontal cortex (dMPFC) was selectively recruited when an adult engaged an infant in a joint attention episode.

However, the results provided by Grossmann, Parise and Friederici (2010) indicate an important difference (lack of right lateralisation) in early infants, as we have seen, but their experiment has maybe the limitation of the stimulus materials represented by static pictures. To overcome such constraints, Lloy-Fox and colleagues (2015) settled their experiments in a more ecological environment. They created two naturalistic social interactions in which *ostensive cues (eye contact and IDS) were presented live and for a long duration*. The researchers used fNIRS techniques for recording cortical responses to the communicative stimuli presented simultaneously. The study was conducted with pairs of infants seated on their parents' lap. Six-month-old infants interacted with a female experimenter who exchanged mutual eye contact with each baby for 15 seconds, meanwhile she uttered «Hungarian nursery rhymes in IDS accompanied by hand movements» (Lloyd-Fox et al. 2015, p. 3). In another experimental condition, she interacted in a one-one communication, in which the combination of the infant direct gaze (IDG) and infant direct speech (IDS) was compared with IDG and the adult direct speech (ADS) modality. Confronting the results of the two conditions, Lloyd-Fox and colleagues found that direct gaze performed by the experimenter «increased neural responses to the multimodal communicative actions (speech plus gestures)» (Lloyd-Fox et al. 2015, p. 7). Therefore, they hypothesised that the detection of direct gaze affected significantly the processing of accompanying communicative signals, i.e. speech and hand movements, recording a strong activation in inferior frontal and temporal regions in both hemispheres. This occurred only when the infants were directly addressed by the experimenter's gaze, otherwise when it was not directed to them or when the infants were performed in ADS modality, they did not process IDS and hand gestures, at least in the brain regions analysed.

The authors invite to remain prudent about the correct interpretation to provide in order to explain neural activation in the cortical areas highlighted by their finding. They doubt that the underlying mechanism is identifiable with TOM, because they did

not find remarkable modulations of the prefrontal cortex (PFC) in response to ostensive cues, as it was noticed in prior works with adults just mentioned above (but see also Enrici et al. 2011), and in infants (Grossman et al. 2008a; 2008b). This fact is likely due to the methodology adopted by Lloyd-Fox and colleagues that is the most recommendable because it reproduces ecological and naturalistic conditions.

However, their results appear to contradict, but only in part, the conclusions provided by Parise and Csibra (2013) to their findings on multimodal ostensive signals. Indeed, Parise and Csibra (2013, p. 2) tested 5-month-olds using ERPs and gamma-band event-related oscillations with static visual stimuli on a computer screen (female face with closed eyes, open eyes with direct gaze, open eyes with averted gaze) combined with IDS/ADS (but using only one word in the two different intonations). They found the same activations in the frontal cortex independently of the presence of one stimulus that elicited an equal response to multiple ostensive signals. This may be in contrast with the peculiar status of eye detection that modulates and enhances the reception of the other ostensive stimuli – as we have seen in previous studies - but we can reasonably argue that this fact could be due to the rapid modalities of stimuli presentation, and a long interaction with the sources of ostensive cues did not occur. Furthermore, Parise and Csibra manifested the same prudence to interpret the nature of mechanism underlying the brain substrates activated during the responses, and they did not speculate about a possible involvement of TOM neural substrates. Rather, as Parise and Csibra (2013, p. 7) underline and Lloyd-Fox and colleagues (2015, p. 12) confirm, the main element that we can figure out in virtue of the statistical results provided consists in what is termed «*obligatory response*» (Parise&Csibra 2013) shown by infants addressed by ostensive communication. In authors' terms: «ostensive signals *obligatorily indicate* to young infants that communication is directed to them [my italics]» (Parise&Csibra 2013, p. 1). Indeed a significant triadic communicative relation must imply not simply an orientation to others' attention, a simple looking at the same thing as another person, rather a mutual *knowing* of looking at the same thing at the same time. Although the first signs of the capacity of shifting attention based on another's change of attention is attested by 4 months of age, when do infants *know* they are attending to the same thing as another person? This is (one of) the challenge of cognitive developmental psychology (Redcay&Saxe

2013; Call&Tomasello 2005). Through infant ERP measures, it has been discovered (Striano&Reid 2006) that 4 month-olds are able i) to detect if a person is looking at an object; ii) process an object if gazed by another; iii) «use another's emotional information to modulate object-processing» (Redcay&Saxe 2013, p. 228).

2.3.2. *Marno and colleagues' study on referential expectation in 4-month-olds*

Once understood that the message is addressed to them, infants are biased to trigger referential expectation that represents a crucial and preparatory moment to approach the informative content of the message. A recent finding seems to highlight the remarkable sensitivity demonstrated by 4-month-olds on the comprehension of the referential nature of human speech in combination with direct eye gaze. This is the case of Marno and colleagues' (2015) study that tested 4-month-old infants to investigate the presence or the absence of referential expectations in infants hearing human speech compared to other auditory stimuli (including silence) in presence of both direct eye contact and object-directed gaze of the speaker. Their striking results highlighted that when infants were looking at a female face uttering a normal speech, they appeared well prepared to find some visual referents of the words, as indicated by their faster orienting towards the visual objects she targeted at the end of the speech (Marno et al. 2015, pp. 1-2). Researchers did not find the same results when the female speaker uttered words in a backward speech or when she opened her mouth and moved her lips without emitting sound. This finding is consistent with the studies mentioned above on the capacity of mutual gaze to elicit referential expectation about an object that becomes the subsequent target of shift gaze-direction. For our purpose, Marno and colleagues' (2015) study is very important for three reasons at least: first of all, they attest (and confirm) that an early joint attention is already available in 4-month-olds. Then, the researchers show how the referential expectation is manifested by the infants, who – third reason - recognise both the *ostensive value of speech* (clearly manifested by expressions such as "Oh! Hello!" and so on), when it is accompanying with other stronger ostensive signals like mutual eye contact, both the *informative value of speech* regardless of the understanding of the meaning (Vouloumanos et al. 2014). Indeed, Marno and colleagues (2015, p. 8) claimed the establishment of the speech referential

interpretation of speech occurring during their experiment after eye-gaze shift of the speaker towards a visual object. Exposed to a combination of ostensive signals (eye contact plus IDS) infants were prepared to seek for potential referents represented by an object and different kinds of speech ("speaker is talking *about* something").

> «The very moment when the speaker averts her gaze to a new direction, the infant will infer that some new and relevant information is being presented to her via the speech signals, and, as a consequence will be ready to seek this information» (Marno et al. 2015, p. 6).

By such cognitive achievement infants are ready to learn from social environment. Therefore, at this point, natural pedagogy may be potentially triggered.

3. TOM and the trigger of natural pedagogy system

The experimental findings shown so far convey a complex framework by which we can figure out some conclusions. From birth, infants discriminate simple ostensive signals with a marked preference for eye detection as evidence about subcortical activations show. Neural areas involved in processing ostensive cues are comparable with adults during ontogenetic development in the first months after birth, but there is not a perfect match under different conditions and modalities of data recording. Therefore, it is better to adopt a cautious explanation that does not predict a primary form of TOM involved in the interpretation of a communicative intention by the young recipients. Here, it is the case to intend the interpretative process as the detection of communicative signals perceived through different channels during a social interaction that can be described as dynamic between an adult and an infant referred to something. The content of the referent is not important at the moment. The point is that infants are able to understand that there is a message for them. It is not necessary to take into account any inferential process about the mental state of the communicator likely expressed in the form: "he wants to tell me something".

An alternative, plausible and more parsimonious hypothesis can be sustained: infants early detect communicative signals normally provided through multimodal

interactions as mutual eye contact and particular vocal intonations. In virtue of such detecting processes, they can comprehend in some circumstances that there is a message for them. Although it is impossible to speak properly about any form of pre-joint attention in newborns, we have seen that also neonates might be able to build dynamic representations of communicative exchanges in terms of attending to some predictable event from the communicative source in front of them (Cecchini et al. 2007; 2011).

Ostensive signals elicit in infants a particular state that is wrong to define as a simple attention, rather it is better described in terms of *referential expectation*, by which if those signals target a concrete visible object present in the contextual environment, infants (at least from 4 months of age) will follow and individuate that object. At this point, ostensive manifestations guide also infants' inferences about the object itself, and the cognitive effects elicited by ostensive communication about generic knowledge are described by natural pedagogy theory. Therefore, we may claim that once the infants are well prepared to pick up the full referential potential of ostensive signals, they are ready to learn about the world from expertise conspecifics, so natural pedagogy system can be triggered, biasing the young addresses of adult communicative intentions to the corresponding informative contents.

For what concerns the preliminary conditions to the comprehension of ostensive cues, I suggest that the "fast-track modulator" model proposed by Senju and Johnson (2009) represents, at the moment, the best explanatory solution posed in the middle between Baron-Cohen's EED hypothesis (1994; 2005), and Csibra's suggestion (Csibra 2010, p. 161) about the presence of an undetermined mechanism able to decode ostension only from perceived visual (above all) and auditory channels. According to Csibra, in fact, the process of decoding ostension through multimodal perceptual channels allows infants to recognise communicative intentions (Csibra 2010, p. 161). However, in order to avoid any ambiguity, it is preferable to adopt another terminology for very young infants involved in the first communicative interactions. It is better to abandon the notion of "*other's intention*", so we can:

i) focus on the minimal early infant comprehension about ostension expressed in the form "there is a message for me", and then, in joint attention processing, "there is a

message for me about something (that object)", rather than the form "he/she wants to communicate a message";

ii) avoid implication in further speculations about TOM mechanism underlying the neural substrates activated, because scientific literature cannot support such hypothesis so far.

I am not claiming that TOM, or its primary form, cannot be involved because of the fact it is not arisen yet. The striking development of contemporary literature about the early onset of mindreading abilities suggests, on the contrary, that a form of metarepresentational TOM starts operating before the first birthday, at least relatively to the attribution of (true and false) beliefs to others (Kovács et al. 2010; Kampis 2017). I will discuss such issue in the fourth and fifth chapter dedicated to mindreading.

The manifestations of communicative intentions pass, pragmatically, through ostensive cues, and in the present chapter, I advance the parsimonious account about the comprehension of ostensive signalling that does not contradict the subsequent TOM commitment. In fact, the neural network involved predisposes the maturation of TOM skills in interactive communication (visual and verbal), as it is clearly attested in adulthood.

4. An innate epistemic principle?

The innate attitude shown from birth in understanding communicative signals may reveal the infants' great predisposition, probably inducted by social-evolutionary pressure, to search for information transmissible by knowledgeable sources. From this point of view we can represent infants as full-blown information seekers, who refine their capacities during the ontogenetic development. However, I wonder whether such predisposition to get information is not restricted to social partners. In other words, what it seems to be crucial in adult-infant transfer knowledge is the ostensive nature of the communicative signals. It would be interesting to test the infants' response to non-social informative sources able both to reproduce clear ostensive signals by technical devices and to provide a referent object. An experiment like this would be important to understand whether human physical features affect so much the comprehension of

ostension that would be restricted to familiar social interactions, or if the recognition of ostensive cues could occur also interacting with unfamiliar artifacts.

Fortunately, Deligianni, Senju, Csibra and Gergely (2011) conducted an experiment with such characteristics, but only with 8-month-old infants, employing non-human objects which were animated on a computer screen and performed distinct motions and sounds for each stimulus induced. They employed a sophisticated interactive eye-tracking technical methodology[13] with a gaze-contingent display in order to use «young infants' spontaneous gaze fixations of the stimulus object» as a kind of response that generated the pattern of contingent reactivity in the object itself (Deligianni et al. 2011, p. 1500). The animated agent on the screen acted automatically without human intervention. Two experimental conditions were created: in the first group (contingent condition), eighteen infants interacted with the animated agent on the computer screen: when the infants fixated the agent in the middle of the screen the latter operated a contingent response. One possible response consisted in a simulation of orientation performed by the agent with a protuberance towards another similar object posed in the left or in the right corner of the screen. A novel response was produced by the agent after a certain quantity of such contingent events. In the second group of infants no interaction occurred with the agent (non contingent condition). Deligianni and colleagues found that 8-months-old human infants followed the animated agent without human features turning towards a target stimulus *only* if the agent had responded several times to them in a contingent interaction. Therefore, the induced contingent reactivity of the unfamiliar animated object on a screen represented a sufficient cue to elicit the infants' orientation-following response. This finding suggests that the «infants' ability to detect contingency and consequent tendency to follow the direction of the responsive agent's subsequent target-oriented behaviour *do not depend on the presence of any other social cues*, such as faces or human voice [my italics]» (Deligianni et al. 2011, p. 1502).

These intriguing results may support the hypothesis that infants are guided by an innate attitude of knowledge request, i.e. an *epistemic principle* that makes them obsessive *information-holder seekers* whatever the source of information is. This attitude reveals a great evolutionary adaptation in the human social world, because the

[13] They used Tobii 1750 Eye Tracker apparatus (Deligianni et al. 2011, p. 1500).

ability to find and detect as much as possible ostensive signals could facilitate and maximise the transmission of knowledge from caregivers and other members of the social environment.

In the following chapter, I will discuss how the natural pedagogy theory describes such knowledge transmission between adults and infants.

CHAPTER 3.

The Natural Pedagogy theory: "A relevance-guided learning device"

1. Precursor studies of natural pedagogy theory

The natural pedagogy theory has been presented to the scientific community by György Gergely and Gergely Csibra in 2005 and in 2006. However, we can claim that to some extent the theory moved its first steps in 2002 after the publication on *Nature* of the article «Rational imitation in preverbal infants» written by György Gergely, Harold Bekkering, and Ildikó Király (Gergely et al. 2002), although, at that time, none of the terms and the notions connected to the theory was used. In that paper the authors replicated - with some novel differences - the Meltzoff's seminal study published in 1988 about an imitation task performed by 14 month-old infants. Andrew Meltzoff investigated the capacity of 14-month-old infants to perform «deferred imitation» of a novel act (Meltzoff 1988, p. 470). Infants dealt with six objects and toys that they had never seen before because they had been built in lab by the experimenters, who did not allow infants to imitate immediately several actions relative to each object, but for the first time only after one week, when the infants turned back into the lab with their parents finding the same experimental setting.

Only one of the six objects[1] had been used by Gergely and colleagues in their replicated experiment, so I dwell upon this one, also because, as pointed out by Meltzoff himself, it was the real novel object that respected some criteria[2]. A wooden box-shaped

[1] The objects/toys were a dumbbell «that could be pulled apart and put back together again»; the second was a flat rectangular base with a thick wooden flap «connected to the base by a hinge. The action demonstrated was to reach out and push the vertical flap over so that it would lie flat on top of the base». The third consisted of a small black box with a button on the top surface: when the demonstrator pushed the bottom a beeping sound was produced. The fourth object was an orange plastic egg: the demonstrator picked it up and shake it. The fifth object consisted in a small stuffed bear that appeared to perform dancing movements on the table. Finally, the sixth object was «a wooden box (19 x 26.7 cm) with a translucent orange plastic panel for a top surface […]. When touched, the panel was automatically illuminated by a light bulb inside the box» (Meltzoff 1988, p. 471).

[2] The object had not been seen, performed or imitated by the infant before; neither it could occur that infants deploy spontaneously the novel object during the play.

lamp had a translucent panel on the top that lightened when touched by experimenter's forehead. The experimenter interacted with the infants calling her/his name, or saying "look over here!". In the familiarisation trial, the demonstration was not so long, and the infant cannot touch or handle the objects. In Meltzoff's experiment infants imitated three of the six acts, but two of them repeated five of six acts. This was a striking result if we consider that each demonstration lasted only twenty seconds. By these data Meltzoff suggested that imitation of novel act occurs in early infants, and that very young children are able to represent the acts that they see in adults and enable to use these representations to guide their own behaviour. Quoting Meltzoff's conclusion: «the behavioral repertoire of infants and their knowledge about objects can expand as a result of *seeing* the actions of others [my italics]» (Meltzoff 1988, p. 475).

What is the nature of such "seeing"? What elements do infants understand, remember, and re-enact after seeing that novel, odd act which is properly an instrumental action? Gergely, Bekkering and Király (2002) started from the consideration that the novel act was a well-structured goal-oriented action, and they wonder «why did Meltzoff's subjects re-enact the head action, when they could just have touched the box with their hands?» (Gergely et al. 2002, p. 755).

1.1 Rational imitation explanation

In order to solve the puzzle, Gergely and colleagues (Gergely et al. 2002) modified the experimental conditions presenting 14-month-old infants two different settings, termed "hand-free" condition and "hand-occupied" condition. In the former, a female demonstrator pretended to be chilly and wrapped a blanket around her shoulders, but soon she liberated her hands and placed them on the table at the two sides of the box. She then touched the box with the forehead turning the light on. In "hand-occupied" condition, another group of infants saw that, after pretending to be chilly, the experimenter kept the blanket on her shoulders holding it with both hands, and, as in the first condition, she turned the light on using her forehead (see also Gergely&Csibra 2006). Keeping in mind the infants' teleological abilities (Csibra&Gergely 1998), Gergely and colleagues tested what action - given "hand-free" or "hand-occupied" constraints - was evaluated by the infants as the most efficient and rational act. They

found that only in "hand-free" condition Meltzoff's results were confirmed: 69% of 13 infants re-enacted the head action, whereas only 21% of the 14 children involved in "hand-occupied" condition imitated the head gesture. Gergely and colleagues concluded that «the early imitation of goal-directed actions is a selective, *inferential process* that involves *evaluation of the rationality* of the means in relation to the constraints of the situation [my italics]» (Gergely et al. 2002, p. 755).

An imitative process cannot explain alone the different performances, also because infants didn't imitate perfectly the head action, as pointed out by Gergely in further papers (infants used sometimes lips, nose, mouth, cheeks and not only forehead). However, invoking the attribution of rationality could be not enough. Would infants evaluate the most efficient act the hand-free head action only because they *saw* that the agent could choose the best action to do? Probably the reason lays in action itself, i.e., the interpretative criterion should be referred only to the fact that the agent is forced to act with her head in hand-occupied. The object's properties were completely obscure to the infants; thus, the infants didn't discover alone the object functions. Rather, the agent showed infants how the lamp worked, and they imitated it freely after one week delay but only the action observed in the hands-free constraint. Likely there ought to be something that guides and facilitates infants to bias the teleological inference given those constraints, something that helps them both to individuate the relevant action-elements to grasp, and to evaluate the instrumental hand-free action performed as the most efficient. A few years later, Gergely and Csibra identified the helpful factor with the notion of *relevance*. They hypothesised that the key element in Gergely and colleagues' experiment (and in Meltzoff's study too) was the familiarisation trial characterised by the mutual communicative exchanges occurred before and during the demonstrations. In other words, the communicative exchanges that preceded the instrumental action and put the infants in a second person perspective would allow them to be guided by relevance, and then to ascribe rationality to the action. The benefit inferred by infants to imitate the unusual gesture was purely pedagogical, i.e., it was interpreted as *an instruction to be learned*.

1.2 Replications of Gergely and colleagues' experiment

In order to test the combination of ostensive cues (as guide for recipients towards the relevant aspects of the referent) and teleological interpretation of an instrumental action, Gergely and colleagues' experiment (2002) has been replicated several times with some variants (different unfamiliar objects and tools and relative demonstrator's constraints) testing either younger infants (age 9-12 months) (Schwier et al. 2006; Zmyi et al. 2009), or those non-human animals able to understand ostensive cues, like dogs (Kupán 2013; Range et al. 2007; Tauzin 2017), or able to understand instrumental action with tools, like apes (Buttelmann et al. 2008). In other words, these additional researches investigated the developmental onset of rational imitation ability from several points of views.

Zmyj and colleagues (2009), for instance, tested three groups of infants *watching video sequences* in which a person illuminated a lamp using the head, while a further control group watched the demonstrator turning on the lamp using the hand. The infants of the three groups could see that «the hands of the model were either free, occupied by voluntarily holding a blanket, or no voluntarily restrained by being *tied to the table* [my italics]» (Zmyj et al. 2009, p. 131). There are two important variants introduced to extend and focus the investigation on the basis of rational imitation: the use of video instead of a live model, and the additional condition of hand tied to table. The researchers expected that «12-month-olds would use their heads to turn on the lamp less often in the condition in which the model's hands were tied to the table (hands-restrained condition), compared to the condition in which the model's hands were free (hands-free condition)» (Zmyj et al. 2009, p. 133). The results were consistent with this expectation, but they did highlight a great difference between 9 month-olds, who performed no significant rate imitation among the four experimental conditions and 12 month-olds performances. The study shows, however, the onset of the capacity to imitate rationally also during such developmental gap. Consequently, Gergely and colleagues' interpretation may apply to 12-month-old infants too, but it is worth noting that «the context-sensitive contrast in imitation was *only* present when comparing the hands-free and hands-restrained conditions [my italics]», because no substantial differences were found between the hands-free and hands-occupied

condition as in Gergely et al.'s (2002) study regarding 14 month-olds. Therefore, this study indicates that «infants are more sensitive to an explicit and non-voluntary contextual constraint [hands-restrained condition] than to an implicit and voluntary contextual constraint [holding a blanket]» (Zmyj et al. 2009, p. 138), and maybe the ontogenetic development promotes a major ability of discrimination.

1.2.1 *Sensitivity to ostensive cues: comparing behaviours between distantly related species*

Some findings on domestic dogs have the scope to test the strength of ostensive cues for social learning among different species. Indeed, «domestic dogs are sensitive to human-given communicative cues and can easily be trained to performed actions that are not causally linked to reward» (Range et al. 2007, p. 868). Comparing similar behaviours and sensitivity in action understanding between distantly related species as dogs and human infants may refer to functional analogies (Topál et al. 2009; Tauzin 2017)[3]. As natural pedagogy theory predicts for infants, the ostensive cues given by the adult/teacher/demonstrator induce in naïve learners a special interpretational stance making them ready to acquire novel information that can also violate the principle of efficiency but that are equally deemed rational and relevant to be learned quickly and permanently. Range and colleagues (2007) designed «an instrumental problem-solving task» comparable to Gergely and colleagues' paradigm, providing the first evidence that dogs imitate «in an inferential selective manner» like children.

[3] There is a flourishing debate in comparative cognition aimed to discover how the very cognition evolves. The emerged problem is that concepts and terminology come often from research field of human psychology, and then are applied to animals. This approach has entailed so far that negative results lead to conclusion that animals fail the test because they really lack an ability, rather because methods are inappropriate. The most striking example in this sense regards the big confusion about TOM abilities in non-human primates (see Chapter 4). For this reason, on the one hand, it is necessary a more explicit and systematic defence of the connection between behavioural flexibility and complex cognition (Mikhalevic at al. 2017). On the other hand, from a neuroscience and biological evolutionary point of view, someone like de Waal and Ferrari (2010) suggests analysing the problem according to a mechanistic, bottom-up perspective including neural underpinnings of cognitive features investigations, molecular-genetics studies (Chittka et al. 2012).

> [Two groups of] dogs watched a demonstrator dog pulling a rod with the paw instead of the preferred mouth action. In the first group, using the "inefficient" action was justified by the model's carrying of a ball in her mouth whereas in the second group, no constraints could explain the demonstrator's choice. […] Dogs imitated the non preferred action only in the second group (Range et al. 2007, p. 868).

Buttelmann and colleagues (2008) compared infants and apes in a selective imitation task confirming Gergely et al.'s (2002) results for 14-month-old infants. Their attempt was to test whether apes were able to understand others' intentions as rational choices of action plans, but the results were unclear, and only a few apes (orangutans) performed results comparable to infants. The experiments would deserve to be replicated maybe without using demonstrators with human aspects.

These findings confirm Gergely and colleagues' (2002) insight about infants' ability around the first birthday to ascribe rationality to well-structured actions observed, i.e., to interpret as rational even those actions that may violate the principle of efficiency. Infants seem to be facilitated, or better, induced to make such inference by encoding ostensive cues, i.e. communicative actions, addressed by the model without the involvement of further mentalising skills (for further experimental evidence see Király et al. 2013).

1.2.2 Paulus and Bekkering's studies

Against such interpretation, Harold Bekkering himself together with Paulus and Vissers (Paulus et al. 2011a) employed five new conditions using the head touch imitation task trying to show why motor resonance mechanism explains the imitation skills in infancy better than the teleological reasoning applied in pedagogical-ostensive context. In their experiment, a female model gave 14-month-old infants the same multimodal ostensive cues across conditions which were analogue to Gergely et al. (2002) experiment with the addition of hands-up condition, button condition (in which «the blanket around the model's shoulders was held by a button so her hands were free» (Paulus et al. 2011a, p. 1053)), and balls-condition (in which the hands of the model were on two balls located on the table). As in Gergely et al. study (2002), the head touch was imitated by more infants in the hands free condition than in the hands

occupied condition, but at the same time few infants imitated head touch in button-condition although they should interpret model's head action as being performed deliberately since her hands were free and not involved in holding the blanket. Also in the balls-condition, following a rational approach we should expect that infants didn't imitate head touch, but in contrast to this prediction, many children imitated model's action because according to Paulus, Vissers and Bekkering (2011a), the perceptual shape of action was very similar to hands free-condition. Infants would not take into account the situational constraints and ostensive cues, they would rather activate a motor program. In this perspective, infant imitation does not depend on rational evaluation, rather on «the similarity between infant's and model's body posture», and the presence of action effects (Zmyi&Buttelmann 2014, p. 21; Paulus et al. 2011 b). These two passages represent the core of the «two-stage model» proposed by Paulus and colleagues (2011b) for infant imitation. It is based on two assumptions: i) the automatic connection between action observation and action execution; ii) the bi-directional association between action and action effect whereby the perception of an action's effect activates automatically a motor program.

Therefore, the experimental results reported by Paulus et al. (2011a) should provide evidence that motor resonance, corresponding to «the matching of another's actions onto one's own behavioural repertoire, is an important mechanism of imitation in infancy», suggesting, furthermore, that «the ability to think rationally about others' actions might not yet be fully functional in early infancy but might emerge and gain importance later during development» (Paulus et al. 2011a, p. 1054-1055).

However, we can quite easily show that the single additional actions demonstrated in Paulus et al.'s findings were not so clearly structured as goal-oriented for 14-month-old infants' perspective, or maybe infants had no time enough to understand and recognise properly the given constraints in order to ascribe efficiency, namely, *to interpret those acts as instrumental actions*. As pointed out also by Buttelmann and Zmyj (2012), the button condition, for instance, requires also the competences to understand the button functions not *necessarily* evident in infants. Similar doubts regard hands-up condition in which infants could not be able to infer what is relevant to be imitated, and also the balls-condition may be interpreted as equal to hands-occupied condition since the balls could lie on the table and not be clearly held by the

performer. Finally, as Paulus and the other authors recognised, their research left open the question whether *imitative behaviour might be susceptible to ostensive cues or to infants' understanding about functional object knowledge*.

To overcome these objections, in a more recent finding Paulus, Hunnius and Bekkering (2013) mounted the head-touch lamp on a rack tilted by 90 degrees. In this way, the infants were able to touch the lamp with the forehead just by leaning forward without putting their hands on the table. According to the authors, infants' imitative behaviour shouldn't reflect any difference between the hands-free and hands-occupied condition since motor resonance was similar in both conditions. However, also in this case, «infants might have viewed the head touch as the most efficient means to illuminate the lamp, because the head was closer to the lamp than the hands in both conditions» (Zmyj&Buttelmann 2014, p. 24).

1.2.3 An integrative model

To solve this apparent contradiction between the two hypothesises, Zmyj and Buttelmann (2014) proposed an integrative model. First of all, they had the merit of highlighting the limits of both approaches:

> «the rational-imitation account lacks a comprehensive theory of how infants transform a visual signal into a motor signal and how infants acquire novel behaviour. […] The two-stage model, on the other hand, lacks empirical evidence that motor resonance and action effects alone play a dominant role in selective imitation» (Zmyj&Buttelmann 2014, p. 24).

Furthermore, Zmyj and Buttelmann (2014) suggest that the two-stage model provides an explanation about *why* infants manage to imitate at all, namely it offers the cognitive mechanism that likely serves infant imitation. In contrast, the rational-imitation account would only explain *why* infants imitate specific actions. *A rational evaluation of actions is conceived as a top–down process since it runs and controls the execution of those actions elicited by motor resonance that is a bottom-up process because it is thought to be an automatic mechanism*. If, and only if, imitation appears to be based on such evaluation, then imitation does occur. «If the model's situational

constraints are not similar to those of the infant, the top–down process inhibits the execution of the action triggered by motor resonance» (Zmyj and Buttelmann 2014, p. 25). It means that the inhibition of imitation is due to infants' inference that the model could not use the most efficient means (her hands) to achieve the goal given those situational constraints.

In conclusion, we may say that the two approaches are not opposite, but they can fill many gaps for the comprehension of infant imitation, and therefore for action understanding and learning mechanism in which the role of social factors is determinant as ostensive cues show by modulating infants' evaluation of the observed action (see in this regard in particular Király et al. 2013). In this sense, we can read natural pedagogy theory as an attempt to provide an explicative model for the cognitive functions involved during imitation processing in a communicative triadic relationship.

1.3 Meltzoff's Like-me hypothesis

Before discussing in details the Csibra and Gergely's theory, it is worth pointing out Meltzoff's account because it has constantly been over time the explicit term of comparison with natural pedagogy for what regards mindreading system's commitment. Indeed, one of the crucial differences between the two approaches consists in the interpretation given to the degree of kinship between imitation and understanding others' mind. In Meltzoff's perspective, imitation and mindreading are *causally* related. Babies have none folk-psychology skill at birth, but they are equipped with «an imitative brain»; the cultural context and social interactions in which infants are immersed together with psychological agents help the early maturation of *an intentional self*, that decodes others' actions through the "like-me" process. *The full-blown mindreading would be the product of such social interplay*.

The notion "like-me" means that infants see, or rather they recognise others like themselves. «*Human* acts are *especially relevant* to infants because they *look like* the infant *feels* himself to be and because they are events that infants can *intend* [my italics]» (Meltzoff 2005, p. 74). Therefore, according to Meltzoff, seeing a human act means to infants recognising an event previously felt in the body. The "like-me" hypothesis suggests that the infants map the other from the Self; in shorthand, infants

recognise when an agent acts as they do: «Through everyday experience infants map the relation between their own bodily states and mental experiences» (Meltzoff 2005, p. 56). In virtue of experiences (action observations) infants are able to project onto others what they felt bodily and registered, but they can do such projection only if they *see equivalence* between their acts and those of others (Meltzoff 2007). How do infants see such equivalence? Infants can build self-other correspondence using their own Self precociously generated by experiences. The "*self-generated experience*" enables to match «their own bodily acts and those observed in others» (Marshall&Meltzoff 2014, p. 2). We shall imagine that infant reasoning is: "your face is similar to my face; my hand is similar to your hand" and so on; thus, infant can imitate human acts because they identify the corresponding body parts. It has been supposed by Meltzoff in many of his works (see in particular Marshall&Meltzoff 2014) that such kind of imitative process is due to human neural mirroring mechanism (Rizzolatti et al. 2001).

According to Meltzoff, infant's imitation and neural representations that underlie it contribute to beget through ontogenetic developmental stages the adult folk-psychology, *but infants do not possess at birth the same mindreading adult framework.* Therefore, also in Meltzoff's view, infants are able to imitate at birth, «but they do not infer intentions from the unsuccessful efforts of others or understand "perception" in others» (Meltzoff 2005, p. 74). The comprehension of intentionality is a later, however precocious achievement. Indeed, one of the crux of Meltzoff's theory is that the putative infant self grows up in the second half-year of life accompanied by a sense of intentional agency that infant uses as a framework for interpreting the intentional actions of others (Meltzoff 2005, p. 75). Briefly, «the self serves the function to understand the actions, goals, and psychological states of others» (Meltzoff 2007a, p. 126).

Such claims imply several assumptions that are not univocally demonstrated and contradict what has been argued in previous chapters about the instrumental and communicative actions. For example, not only human acts are relevant for infants, and they understand even unfamiliar and physically impossible events. This should be hard to explain only with self-others correspondence based on prior experiences. Moreover, if infants understand events that look like how they feel, then they have probably access to some kind of introspection that originates in the body first of all. The projection of

their own selves to others entails a metarepresentation, namely, a conceptualisation of those body feelings, but we don't know how so young children could be self-conscious. Putting it in a nutshell, such as *feeling one's* own self implies a *strong first-person* based on experiences, and in particular on the early observation of bodily acts of others.

Considering the early onset of such overall attested capacity, i.e. the intentional imitation of others' behaviour, a very simple question arises: how can contingent and individual experiences (accomplished in social parental context) generate infants' competence in action understanding? Individual and contingent experience should regularise the whole process, so we would expect from the experimental samples a major differentiation in the early stages of ontogenetic development depending on the subjects' social and familiar experiences, but this has not been discovered, rather the opposite case.

Another critical point assumed by Meltzoff, and still much debated, regards the role of mirror neurons (MN) and their supposed conceptual nature[4]. In brief, the very activation of MNs due to the observed actions would determine the recognition of others' motor intentions by a replication, rather a simulation (without execution) of their motor behaviour. The term "mirroring" refers to such sensorimotor simulation. In general, we may say that MNs request a rigid computation since they depend on perceptual inputs and automatically trigger a simulative response. More precisely, MNs should serve as the mechanism underlying the human capacity to "copy" observed novel actions. In this regard, Rizzolatti and Craighero (2004) hypothesise that the human MNs generates a new motor "copy" from an observed human action through a "decomposition-recombination" mechanism:

> «During learning of new motor patterns by imitation the observed actions are decomposed into elementary motor acts that activate, via mirror mechanism, the corresponding motor representations […]. Once these motor representations are activated, they are recombined, according to the observed model by the prefrontal cortex» (Rizzolatti&Craighero 2004, pp. 182–183).

However, in order to provide an explanation for Meltzoff's imitation task, such «simple notion of direct resonance between observation and execution is not sufficient»

[4] For a review about this issue see Meini&Paternoster 2012.

(Marshall&Meltzoff 2014, p. 2), as Meltzoff himself noticed. It is necessary a representation of observed events that can be retrieved to perform a matching re-enact at a later time, and motor system involvement is not sufficient for explaining this re-identification processing.

1.4 Conceptualising MNs: hypothesises

Pierre Jacob (2008; 2009a; 2009b) sustained a conceptualist interpretation according to which MNs realise "motor concepts". In his view, MNs are not merely resonators, they rather work as symbols that stand for classes of actions, exhibiting at least some of the typical features of concepts: «abstraction, integration, multimodality, and inferential roles» (Meini&Paternoster 2012, p. 197).

The conceptual nature of MNs proposed by Jacob is radically different from the interpretation provided by Gallese and Lakoff (2005), who claimed that the functional roles attributed to concepts can be held by sensorimotor representations, or in other words, *all* the concepts (even abstract ones) can be identified with sensorimotor representations. In their view, motor simulation is necessary and sufficient to categorise and make inferences, understand language and read others' minds..

A more cautious account has been advanced by Meini and Paternoster (2012), who proposed to intend MNs' activity as conceptual mechanism only if «motor concepts are partly realised through the direct activation of (pre-)motor areas» (Meini&Peternoster 2012, p. 198). In other words, understanding the concept "*grasp*" would necessitate simulating the act of grasping through the MNs' activation.

Even if we assume the conceptual nature of MNs, are we legitimated to attribute to MNs the underlying mechanism to imitation in infancy?

1.5 MNs and infant imitation

For what concerns the alleged constitutive role of MNs in infant imitation we remain in the field of hypothesises. What experimental findings have attested so far, applying EEG methods to early infants, is the presence of changes in sensorimotor alpha (or mu) rhythm[5] although it occupies a lower frequency than in older children and

[5] For more details see Chapter 1, § 4.2.

adults (Marshall&Meltzoff 2014). Such neural activity changes are associated with «execution and perception of biologically meaningful stimuli» (Vanderwert et al. 2013, p. 16). In brief, the neurobiological mirroring is associated with changes in the mu rhythm. Quoting Marshall and Meltzoff (2014, p. 2), it has been acknowledged that «infants have a bidirectional maps between action perception and their own action production», whereby during action's observation an *internal motor representation* of that same behaviour would be activated within the young observer (Hunnius&Bekkering 2014, p. 2). However, in my opinion, by referring to MNs mechanism the "like-me" hypothesis doesn't manage to fill the evident psychological gap between the bidirectional process whereby visual input generates a matching response and vice versa, and the fast maturation of a strong intentional self that would be originated in few months of life by the cluster of bodily feelings, and that would be able to project one's own intentions to others.

An epistemological puzzle remains closely connected to this issue and it consists in the risk of circularity in the argumentation of action understanding starting from MNs mechanism. As I sketched in the first chapter, even if the sensorimotor activation occurs in action understanding, the problem is to discriminate the causal direction. Put the case that 1) we reject the reductive Csibra's interpretation of neural mirroring (2007a, p. 454) –according to which MNs is only «a direct, unmediated, automatic, mandatory, resonance-like transfer mechanism» of single movements -; and 2) we claim, following Sinigaglia (2010), that the neural resonance regards the action structure, the problem is likely to ground the hypothesis that MNs *causes the comprehension of that structure*, for instance, goal-oriented actions. What evidence tells us is a correlation between imitation (observation-production) of acts and the activity of MNs, *but why does the degree of such activity change depending on the different kinds of action?* As Pomiechowska and Csibra (2017) pointed out, the different degree of sensorimotor activation between instrumental and communicative actions may indicate a top-down processing of action understanding that determines *causally* the trigger of MNs (see Chapter 1), and not the opposite.

In my opinion, this evidence entails changing the framework of triadic relation between an agent, an observer and the referent from the action-oriented paradigm, since *the comprehension of agent's behaviour is primarily tied not to the agent herself,*

intended as the target of the observer's intentional projections, rather to the structure of intentional action performed and its relative context with environmental constraints. Motor (or mirror) system is not always recruited for understanding and predicting actions, i.e. for inferring false beliefs, as Southgate and Vernetti (2014, p. 9) showed in their experiment; rather it is involved only in some cases. This may induce to think that the primary action elaboration is made elsewhere by other kinds of networks and cognitive systems, and thus motor activation occurs after a prior form of action understanding.

1.6 Tomasello et al.'s theory of infant social learning

Tomasello, like Meltzoff (1999, p. 82), suggests that imitative learning relies on such infants' identification tendency. However, for Tomasello and colleagues (2005) identification has not to refer only to the level of observed motor behaviours, but to the human specific primary «motivation to share psychological states of others» (p. 1) that leads to «more deeply psychological levels of identification» (p. 26). In virtue of such simulation and identification mechanisms that occur at mentalistic level, the infants can understand how others' behaviour is designed to achieve the goal (Tomasello 1996, p. 324). This entails what Gergely and Csibra (2005) have named the «cognitive transparency criterion», whereby infants access to other's mental intentions and the reasons behind others' «rational choice of behavioral strategy», that is supposed to serve as *the relevance selection filter* determining «which aspects of the behaviour are relevant for reproduction» (Tomasello 1996, p.323; see also Tomasello et al. 1993). According to Gergely (2007a, p. 184), both explicative strategies are not able to explain Gergely et al. (2002) experimental results. In fact, given the assumptions relative to «mechanisms of direct action mapping and identification-induced drive to imitate human actions», Meltzoff's and Tomasello's models should expect analogue amounts of imitation induced both in hands-free and hands-occupied conditions. Instead, natural pedagogy enables to take into account the different infants' imitative behaviours. In shorthand, human infants are able to learn rapidly cultural relevant information *from* others through pedagogical knowledge transfer permitted by ostensive communication, rather than learning such information «through the adult's perspective in a truly

intersubjective fashion», as proposed by Kruger and Tomasello (1996, p. 371) (Gergely 2007a, p. 185). In order to provide the theoretical grounding of infant imitation, intended here as central mechanism responsible for intergenerational transmission of human cultural forms, it is the case to turn to natural pedagogy account (see § 2) that embraces an action-oriented conception of self-others (i.e. infants/adults) relation, whereby the others are seen as important sources of knowledge. In natural pedagogy framework, the others represent potential cognitive treasure-bearers committed to transfer social and cultural contents toward particularly sensitive naïve recipients.

2. Natural pedagogy theory

From birth human beings are immerged in a world of social norms that regulate behaviour and involve the use of objects, whose functions are not immediately cognitively transparent, like, for instance, eating a plate of soup with the spoon. Most often the types of cultural knowledge contents have no adaptive value, they are "simply" arbitrary and conventional. However, cultural forms dominate the life of social community and they are preserved and transmitted over time. Such characteristics of human life style emerged thousands of years ago, therefore a learnability problem connected to the transmission of know-how and social conventions arose. Gergely and Csibra argue that during hominid evolution natural pedagogy has been selected to ensure fast and efficient acquiring and intergenerational transfer of cultural knowledge overcoming the hard social environmental conditions of «cognitive opacity» of human cultural forms (Gergely&Csibra 2006; Csibra&Gergely 2006; Gergely 2007a; Gergely 2013). Sometimes Gergely and Csibra have defined human pedagogy as a «Mother Nature's trick», because of the fact that in order to guarantee an optimal transmission of cognitively opaque relevant knowledge, human beings developed a complex «dedicated communicative system» in which, on the one hand, *knowledgeable individuals are naturally inclined to provide manifestly their cultural baggage to naïve conspecifics*, whereas, on the other hand, the latter are equipped to receive and assimilate the information transferred by *searching and attending* to such (communicative) manifestations (Gergely&Csibra 2006). The modalities and the means by which these manifestations occur represent the hinges of natural pedagogy system. In this regard Gergely (2007a, p. 170) claimed:

> «Pedagogical knowledge transfer is triggered by specific ostensive and referential cues to which infants show special and early sensitivity and which involve the selective communicative manifestation of relevant and generalizable cultural knowledge about referent kinds for the infant to fast-learn».

As suggested by Gergely et al. (2007) pedagogical cues are as an "interpretation switch", signalling to infant that she is being taught and receiving relevant information. For this reason Gergely (2007a, p. 173) defined natural pedagogy as «*a relevance-guided social communicative learning device* of mutual design [my italics]» able to guarantee effective transfer of relevant cultural knowledge through multimodal ostensive communication. The latter predisposes infant shaping her inferential processes about epistemic information along two reasoning pathways (or biases): one directed towards the object categorisation, and the other one towards the adults and social context. The first inferential pathway is named «assumption of generalizability», whereas the second pathway is termed «assumption of universality» (Gergely et al. 2007, p. 141). Through the former, infants learn not just episodic and local facts «but the generic structure of their cultural worlds» (Tomasello 2016, p. 643). Such bias leads infants to infer that «the pedagogically manifested information about the referent is generalizable to the object kind that the referent belongs to» (Gergely 2007a, p. 179), whereas through the latter, infants consider the epistemic information as shared with other members of the community, i.e., as already known by everyone. The universality assumption can be expressed by the following sentence: «If someone knows something, everyone knows it» (Csibra&Gergely 2006, p. 273). In other words, such assumption triggers the bias whereby what is just learned by infants is alleged by them to be already known by everyone. Quoting Gergely, the universality bias elicits «the implicit expectation by the infant that the manifested information will contain publicly shared universal cultural knowledge available to all others (and not only to the demonstrator, who is the communicative source of the information)» (Gergely 2007a, p. 179).

I deem that the universality assumption is crucial for the efficacy of transmission and above all for the maintenance of cultural knowledge under the form of conventions and common ground beliefs between generations. Such implication of natural pedagogy would represent a fundamental strategy for cultural transmission enabling to optimize the reinforcement of the social practices, beliefs, and values of communities.

Everywhere «children draw on a repertoire of cultural learning strategies that optimize their participation in and acquisition of the particular practices, beliefs, and values of their community» (Legare&Harris 2016, p. 633; see also Legare&Nielsen 2015). In fact, learning from others implies not only the acquisition of practical competences but also norms, rituals and beliefs of the own cultural group. In this regard, Tomasello acknowledged that in virtue of natural pedagogy «human children do not just culturally learn useful instrumental activities and information, *they conform to the normative expectations of the cultural group* and even contribute themselves to the creation of such normative expectations» [my italics] (Tomasello 2016, p. 643).

2.1 Functions of ostensive cues within the natural pedagogy theory

The comprehension of the referential nature of ostensive cues allows natural pedagogy system to start working; in fact, ostensive cues addressed to the infant learner would trigger:

i) The «automatic interpretation» that the holder-source knowledge has a referential intention. It is as if the infant thought: "I am going to be 'taught' something new and relevant!" (Gergely 2007a, p. 178).

ii) An implicit expectation as if infant thought: "What am I going to be 'taught' about?" In other words, ostensive cues facilitate the identification of the referent about which infants expect relevant and new knowledge to be manifested.

iii) The «presumption of relevance» whereby the naïve learner assumes the manifested information as *relevant and reliable*. It entails that when infants are learning novel behaviour, they don't have to evaluate the importance of the reason why a person performed that action. Infants do not need to infer the other's mental state underlying the action observed. In most cases infants might not comprehend the particular reason for an action, but due to *epistemic trust*, they flexibly adapt their effort to acquire new informative contents and skills. The evaluation is made referring to the action itself.

iv) Two further implicit assumptions that constrain the infant's interpretation of information content: *generalizability assumption*, and *universality assumption*.

The operative functions of natural pedagogy system are iii), iv) e v), which I am going to show in details.

2.2 The epistemic trust *triggered by ostensive communication*

The epistemic trust is a crucial component of pedagogical stance because it allows considering others as benevolent and reliable sources of cultural information, activating the *presumption of relevance* about manifested knowledge contents without needing to test them or critically evaluate their validity any further. In this way, beliefs in the form of cultural contents can be transmitted and accepted in virtue of the authority of the source; Recanati (1997) named as «*deferential* such culturally transmitted beliefs» (Jacob&Gergely 2012, p. 60). I will unfold these concepts in Chapter 6. Now it is important to underline that *deferential beliefs are based on trust*, and their acceptance depends on the degree of *relevance* communicated by the informative intention of an agent (Sperber&Wilson 1995), even if they are not fully understood by the recipient (Recanati 1997). Because of the bond to relevance, such beliefs are not entirely unjustified, they are *semantically determinate*, but (in virtue of their opacity) *epistemically indeterminate*. If an action observed by infants is confused, for instance, it cannot be represented any way, like a sentence that contains uninterpreted symbols – Recanati claims – «cannot make its way into the mind (whether into the belief box or elsewhere)» (Recanati 1997, p. 91). However, if an action involving an object function is clear and well structured, the infant is able to represent it (its content will be relative to an executive function), even if the infant ignores a lot of important and meaningful aspects connected to the objects and/or to the background of the action itself. Paraphrasing Recanati's terms, infants may not be aware of the propositional contents of the deferential representations that they come to accept, but these kinds of representations have the *proper form of a belief*, whereas 'belief' is to intended here as whatever representation is stored in the so-called *belief box* «in such a way that it can be

freely used in practical or theoretical inference» (Recanati 1997, pp. 94, 98). I will provide further details about this issue in Chapter 5, anyway it is worth anticipating that infants' information process has the form of a propositional content, i.e. of a belief intended as knowledge contents semantically determined but eventually also epistemically undetermined. This is the price to be paid for rapidity of acquisition compared with the lack of evaluation capacities. Nevertheless, the information contents have and maintain propositional formats that are stored and attributed to others.

In the following paragraphs, my attempt is to show how this notion of belief as informative content[6] is managed by infants in pedagogical context. It is important to underline that, within the theoretical framework of natural pedagogy, epistemic trust is the fundamental component allowing a faster achievement of the manifested information, guaranteeing the young learner to fix this knowledge content. The very fact that the infant has formed such belief, i.e., the fact that representation of such content is stored in her belief box, constitutes a sufficient justification for persisting to hold it, and ascribe it in the forms we will see below, even in the presence of counter-evidence. However, as suggested by Gergely and colleagues (2007), a crucial ontogenetic developmental acquisition is to mitigate the strength of such deferential attitude induced by pedagogical cues. In fact, deferential attitude has to change and develops in favour of what Sperber and colleagues (2010) have defined *epistemic vigilance*, «that assess the quality of incoming information and the trustworthiness of the individual who dispenses it» (Mazzarella 2016, p. 183). Sperber and colleagues (2010) sustain that humans have developed such cognitive alert against the risk of being misleading and deceiving by social interlocutors[7]. As already suggested by Koenig and Harris (2007) and confirmed by Sperber et al. (2010), such vigilance is both directed at the content of knowledge and at the communicator, whereby «it requires an understanding not only of a communicator's epistemic states but also of her intentions, including intentions to induce false beliefs in her audience» (Sperber et al. 2010, p. 18); reasoning

[6] I reserve to specify the notion of belief in the next two chapters.

[7] According to Mazzarella (2016), the epistemic trust not only affects the reliability and the believability of the communicated information, but it contributes, interacting together with other cognitive mechanism underlying comprehension processes, to the acceptability of interpretative hypotheses about the speaker's meaning.

competencies and background knowledge enable to confirm, or dis-confirm, the validity of speakers' utterance[8]. Mascaro and Sperber (2009, p. 367) suggested that the achievement of an accurate and sophisticated epistemic vigilance proceeds by degrees throughout development.

They investigated vigilance toward deception in 3- to-5-year-old children using three studies, which showed that younger children prefer «the testimony of a benevolent rather than of a malevolent communicator», while only from the age of four children seem to comprehend «the falsity of a lie uttered by a communicator described as a liar» (Mascaro&Sperber 2009, p. 367). Finally, starting from the fifth-sixth year of life children use in the field the ability to recognise a lie when they are warned that the communicator intends to deceive them. On the basis of these findings, Mascaro and Sperber claim that epistemic vigilance emerges along such developmentally and functionally distinct three step-components: the preference for the testimony provided by a benevolent communicator, the understanding of the epistemic values of a deception (the fact that it is a lie), and finally the understanding of the intentional aspects of a deception.

Furthermore, Bas and Sebastián-Gallés (2016, p. 62) investigated how *dominance status* helps infants determine the most reliable informant between different adults holding conflicting knowledge. They found that «towards the end of the second year of life, children can use information about dominance status to guide their learning».

All these findings demarcate, as I will argue in Chapter 6, the temporal limits of the natural pedagogy system operational scopes and its social communicative contextual domains. However, the course of natural pedagogy cognitive activity along with the developmental stages are far to be fixed, and much more is still to be discovered.

[8] In this regard, see also Birsch et al. (2008, p. 1018) who found that «3- and 4-year-olds favor a previously accurate individual when learning new words and learning new object functions and applied», in respect of inaccurate individual. In other words, infants keep track spontaneously of an individual's history, and «use it to guide subsequent learning without any prompting».

2.3. *Assumption of generalizability*

Natural pedagogy hypothesis predicts that ostensive signals orient infants to categorise the referent object according to its kind. *Generalizability* means that the object functions learned in pedagogical context are generalised into a proper kind, in order to be used in other future contexts. In this way, what infants are learning does not only regard the local and single object, but the categories which objects belong to. Several studies indicate how infants (at around the first birthday) assume that an object cannot belong to two different basic-level kinds, i.e., a duck is not a ball not only for the different shape or colours, rather because 'duck' and 'ball' are (and belong to) two different types of entities, that have specific features and properties.

Conceptual categorisation is grounded on object-functional properties rather than on perceptual similarities between class exemplars. As Mandler (2000) claimed, conceptual categorisation creates the notion of kinds (for instance, 'humans', 'plants', 'vehicles', 'animals' and so on)[9]. Categorisation is closely tied to language processes: words may sharpen category boundaries, and enhance category inductions beyond external similarities, shift attention to relevant properties of objects. In fact, it has been pointed out that verbal labelling facilitates encoding objects in terms of kind because the verbal label can represent an «essence placeholder» for the kind (Xu 2005, p. 85; see also Hall et al. 2008; Xu 2002; Xu et al. 2005; Gliga et al. 2010). However, verbal labelling is not necessarily the unique causal factor to induce kind-based representations for objects. From very early on, infants pay attention to function-correlated properties of objects connected to the comprehension of physical events (Baillargeon et al. 2012), and verbal information may support such infants' sensitivity to artifact functions. Futó and colleagues (2010) hypothesised that communicative (even non-verbal) demonstrations of artifact function are sufficient to create kind-based object representation (at least in 10-month-olds), as predicted by the generalizability bias of natural pedagogy. The researchers started from the consideration that a «demonstration of artifact function is a

[9] Here, with "conceptual categorisation" I mean, quoting Nelson and Snyder, a «conceptual information […] thought to be acquired by attentive, conscious analysis and a "redescription" of perceptual information into conceptual form» (Nelson&Snyder 2005, p. 3).

goal-directed action that involves a causal intervention on the artifact by an agent, which manifests a specific dispositional property of the object» (Futó et al. 2010, p. 2).

Hernik and Csibra (2009, p. 34) showed how young human infants are very sensitive to functionally relevant features of objects, learning tool use and inferring tool functions «from others' goal-directed actions and demonstrations». Starting from this point, Futó and colleagues (2010) assumed that «the kind-specific function of an artifact can only be inferred from its potentially relevant physical properties, its re-occurring instrumental use observed in specific types of goal-directed activities, or its observed culturally conventional use» (Futó et al. 2010, p. 5). Furthermore, they noticed that in prior studies (focused on kind-based object individuation) (e.g., Xu&Carey 1996; Xu et al. 2005) ostensive signals such as eye contact and IDS were always present, but they were parts of experimental conditions not sufficiently highlighted.

2.3.1 Futó, Téglás, Csibra, and Gergely's experiment

In particular, Futó, Téglás, Csibra, and Gergely (2010) investigated the ability of 10-month-old infants to represent objects in terms of their kinds aiming to verify «whether communicative ostensive signals could play the same role as verbal labelling in enabling 10-month olds to rely on property information in an object-individuation task» (Jacob&Gergely 2012, p. 77-78). In their experiment infants were familiarised to see two novel objects created in the lab by the researchers: one was a red rectangular box and the other one was a pink bell-shaped box. The former had a circular dial on its upper-mid portion: if the dial turned, a short musical melody played. The latter had three lights, red, orange and green, and a handle on the top: pulling the handle all three lights flashed simultaneously. The objects did emerge one at a time from behind a screen, and when at a certain point the screen had been removed, infants could either see the two objects or only one. In the first experiment, two conditions were settled: in the Communicative Function Demonstration (CFD) condition the functional uses of the novel objects were shown sequentially from behind an occluder; whereas in the Baseline Condition (BC) infants were presented «with only the stationary displays of one or two objects» (Futó et al. 2010, p. 3). In CFD condition, while the artifacts were hidden behind the screen, 10-month-olds were ostensively greeted by IDS (with a

female Hungarian voice saying: “Hi baby, hi!”). During the familiarisation trials, infants (sitting on their parents’ lap in front of the screen) saw a hand pulling out (from behind the screen) an object that followed a horizontal trajectory until it stopped and stayed, while the hand performed the function demonstration twice manipulating or the dial or the handle. Finally, the hand pulled back the objects behind the screen following the same trajectory. During the interval between the function demonstrations, when no object was visible, another ostensive cue (IDS) was provided. After the familiarisation trials, the infants received two test trials: the first one was identical to the familiarisation trials, while during the second one infants saw the hand that removed the screen revealing either both of the objects or only one. The measures of infant looking time revealed how the babies looked reliably longer when only one object was shown rather than two. However, a second experiment was necessary to test whether the effect of function demonstration on object individuation was dependent on IDS ostensive cue or on hand intervention on the object. For this reason, the researchers created a Non-Ostensive Presentation Condition (NOP), and Non-Causal Intervention Condition (NCI). In the first experimental condition, the IDS greetings were replaced with «non-human melodic sound generated to match the surface acoustic parameters of the original ostensive stimuli» (Futó et al. 2010, p. 5). In NCI condition, the objects moved by themselves displaying the same movements of the Experiment 1, and the actions were preceded by ostensive cues like in the Experiment 1. This time, however, infants looked for longer (or quite approximately equal to the rest) the scenario where the two objects were present simultaneously in NCI. The lack of remarkable difference between the two conditions and scenarios seem to indicate that 10-month-olds require both the presence of ostensive signals and the presence of manual causal intervention to accomplish object individuation grounded on artifact functions. Therefore, infants observe an instrumental action, i.e. the «causal agent’s use of an artifact to achieve a specific effect», but to assign a kind-based representation to the observation of functional use, infants need to be ostensively informed. In this way, infants interpret the whole action as «a communicative demonstration to manifest the kind specific [object] function» (Futó et al. 2010, p. 6).

Due to these experiments we can ascribe to ostensive signals (IDS exactly) only a facilitating role for object individuation. In order to better clarify this interpretation, a

third experiment has been conceived, where the two different functional uses (i.e., the handle and the dial and their relative effects) were demonstrated on a single novel object. «This object was created by fusing the two artifacts in Experiment 1 into one object that contained all the function-relevant features»: the object was akin to pink bell-shaped box with the circular dial attached to the middle of the pink object (Futó et al. 2010, p. 6). Only such object was used in the familiarisation and the first phase of the test trials that occurred identically to CFD condition of Experiment 1. However, during the second phase of the test trial the occluder was lifted up from the screen, and the infants could see an identical object they had only seen in the familiarisation and at the beginning of the test; or two novel objects never seen before: a pink bell-shaped with a handle and three lights, but without the dial, and 2) the pink bell-shaped box with the dial but without the lights and the handle. The researchers termed such experimental schema "Double-Function Demonstration Condition" (DFD) that was compared with Baseline Condition as in Experiment 1, and the results were perfectly equivalent to the first experiment. Although the two functions were demonstrated on the same objects, «infants did not encode the object during the familiarisation events simply in terms of its overall visual features» (Futó et al. 2010, p. 7). Indeed they looked for longer the single object identical to the familiarisation trial than the two distinct objects, as if they assumed that «only two distinct kinds could exhibit two distinct functional properties» (Jacob&Gergely 2012, p. 79). Futó and colleagues (2010) argued that communicative demonstration of a single functional property of the double-function object, provided at a given time, biased the infants to infer the presence of two objects rather than only one. In other words, according to the authors, infants assume that «basic-level artifact kinds are defined by a single essential function» and such assumption «*produced the illusionary inference to the presence of two objects behind the occluder*» [my italics] (Futó et al. 2010, p. 7).

What we can figure out from these three experiments is the fact that if types of information as visual features, emotional valence, tactile properties may contribute to recognition of objects, they do not manage to determine their kind membership. Furthermore, if the simple observation of object's use is not sufficient for 10-month-olds to interpret such information as the proper function of the artifact (as Experiment 2 shows), on the contrary within an ostensive communicative context, infants "read" «the

manifested function as indicative of an artifact kind» (Futó et al. 2010, p. 8). The conclusion is that non-verbal demonstration of artifacts (i.e. ostensive communication, IDS in this case) enables infants (already at 10 months of age) to have kind-based object representation in the very same way as linguistic labelling can do (as illustrated by Xu 2002). In a few words, the identification of artifact function is well served by the «communicative demonstration of its kind-specific functional use» (Futó et al. 2010, p. 5). Following the results of Futó and colleagues' experiments, we may claim that infants use artifact functions as *indicators of kind membership*; furthermore, the third experiment's interpretation highlights that the infants expect that one specific function defines one specific kind. Namely, it seems to be «one-to-one mapping between functions and artifacts» (Futó et al. 2010, p. 8)[10].

2.3.2 Generalizability in preschoolers: Butler and Markman's experiment

To sum up, we have seen that through ostensive communication infants may infer both that the information being communicated is relevant (as predicted by Sperber and Wilson (1986/1995)), and more specifically that the information transferred ostensively is *kind-relevant* and *generalizable*. Such process begins very precociously, and it deserves particular attention understanding how the generalizability bias develops through the first years of life for analysing the reliability and the limits of natural pedagogy system[11]. Butler and Markman (2012) investigated preschool-aged children (3-4 year-old) during exploratory play with particular objects. Some of these novel toys were inert and others had special properties, but they were all identical. The authors

[10] These claims suggest an *essentialist* construal of artifact kinds, as also pointed out by Jacob and Gergely (2012). Such suggestion may have important entailments for the notion of *psychological essentialism* in infancy, according to which categories as 'boy', 'girl', 'lion', 'tree', 'ethnic group', etc. have an underlying invisible reality, and causal properties not directly observable that someway gives the category (and all of the members) its identity. See §§ 7 and 7.1 below for more details.

[11] Another important experiment conducted on 10-month-old infants is the Topál et al.'s study (2008) that reproduce the Piagetian "A-not-B error" task. The researchers applied natural pedagogy theory (and generalizability assumption in particular) to this kind of task to interpret under another theoretical framework the perseverative errors made by 10-month-olds in hiding and search game with an object and two containers. I will discuss this experiment in Chapter 6, § 5.2.

tested whether the children would have formed «different expectations about generalizability depending on whether a novel property was demonstrated pedagogically or produced in a non-pedagogical manner» (Butler&Markman 2012, p. 1417). Starting from the fact that 4-year-old children, and even younger, take linguistic labels as referring to kinds, Cimpian and Markman (2009, p. 14; and see also Cimpian&Cadena 2010[12]), claimed that information conveyed generically (for instance expressed by the sentence: ‘‘snakes have holes in their teeth’’) influences strongly children's kind representations and it «is more conceptually central» than the information conveyed non-generically (e.g., by the sentence "this snake has holes in his teeth").

In this regard, Butler and Markman (2012) hypothesise that preschoolers are able to process information received pedagogically in a more generalizable and conceptually central manner, exactly as they do for information communicated generically. In their experiment, the novel objects presented to thirty-four 3-year-olds and thirty-four 4-year-olds were *eleven small wooden blocks*, and the only *active* block had thick magnetic tape on one end; the other *inert* blocks had identical but nonmagnetic tape on the end. The experimenter taught *a novel linguistic label for the magnetic block* "blicket". After the presentation, the experimenter used a distracter play for a few minutes and then the test trials started under two conditions. In the pedagogical condition, when he picked up the *active* blicket he said: ‘‘Look, watch this!’’ and he *deliberately* began to place the blicket on the paperclips, while in the accidental condition, he seemed to «*accidentally* drop the blicket on the paperclips as he was putting it away, exclaiming ‘‘Oops!’’» (Butler&Markman 2012, p. 1418). Finally, he located the ten inert blocks (that lack the magnetic property) on the table, inviting verbally the children to ‘‘go ahead and play’’, whereas he sat facing away from the children for one minute leaving the table.

Researchers were interested in the type of exploration children performed looking for the magnetic properties of the ten inert blocks, or other kinds of function the blocks could have. For this reason, three aspects of exploration were investigated: the *time* spent to explore the inert blickets, «the *number of attempts* children made to elicit the

[12] Cimpian and Markman (2009) tested 4 and 5-year-old children, and Cimpian and Cadena (2010) too but they used the generic example: ‘‘Dunkels are sticky") compared with the non-generic sentence ‘‘This dunkel is sticky".

property from the inert blickets», and finally the *number of inert blocks* explored by the children (Butler&Markman 2012, p. 1419). The researchers found that in the accidental condition the younger children did not explore the inert blickets at all. «This suggests that 3-year-olds in the accidental condition may have failed to notice the property, or might have been hesitant to engage in exploratory play», while in pedagogical condition all the children «*persisted* in trying to elicit the property from» inert blocks (Butler&Markman 2012, p. 1420), and 4-year-olds spent significantly much more time attempting to pick up unsuccessfully paperclips with the inert blickets in pedagogical condition (as the younger children) rather than in accidental condition[13]. The authors interpreted the continuing exploration of few inert blickets shown by the children in the pedagogical condition as strong evidence that the children understood the property as generalizable information about the kind, inferring that even the few kind members they stressed should share the same property.

Furthermore, in a second experiment, Butler and Markman (2012) tried to encourage children exploration adding another pedagogical cue, i.e. an enthusiastic experimenter's reaction ("Wow!") uttered after the property demonstration of the active blicket in both conditions. This time some 3-year-olds tried at least one inert object in both conditions, and as in the prior experiment, both children groups «made stronger inferences about the generalizability of the novel property when it was pedagogically demonstrated, as measured by their exploration of the inert blicket» (Butler&Markman 2012, p. 1421). However, I would like to shift the attention towards a third experiment where the researchers added the *intentional* condition in which the blicket property was demonstrated voluntarily by the experimenter, but in the absence of any pedagogical cues. The procedure was the same of the second experiment, but in the *intentional* condition, «after putting away the distractor items, the experimenter picked up the active blicket and placed it deliberately on the pile of paperclips» without making eye contact or establishing joint attention with the children (Butler&Markman 2012, p. 1423). The results indicated that when the experimenter used intentionally the blicket for the novel function without pedagogical cues, 4-year-old children made «weaker inferences about its generalizability» spending less time in the exploration of inert objects. On the contrary, 4-year-olds appeared to make stronger inferences about

[13] The study reported the results from each age group separately.

generalizability when the property was demonstrated pedagogically rather than it was manifested accidentally or intentionally without ostensive cues. This was evident through the time spent and the attempts accomplished for exploring the functions of inert blocks. In conclusion, we may claim that by age 4 «children modulate the strength of their inferences about the kind-relevance and generalizability of a novel property on the basis of whether it was demonstrated pedagogically for their benefit» (Butler&Markman 2012, p. 1425).

For what concerns 3-year-olds, the data confirm a quite different behaviour compared to older children. Indeed, 3-year-olds spent more time exploring the inert objects in both the pedagogical condition and the intentional condition, rather than in the accidental condition. This could mean that they are able «to make a strong inference about generalizability simply on the basis of seeing a novel object intentionally used for a particular function» (Butler&Markman 2012, p. 1425). Therefore, if 4-year-olds are able to better differentiate intentional actions from pedagogical ones, 3-year-olds base kind membership generalisation regardless of whether the object demonstration is done intentionally, or the object-function is manifested for their benefit. In conclusion, we may claim that when a causal property is explicitly manifested pedagogically, children make strong inferences that drive and condition their explorations about object-functions despite counterevidence. This evidence supports the authors' hypothesis whereby pedagogical demonstration «conveys information about the generalizability and conceptual importance of new information» (Butler&Markman 2012, p. 1425) as generic language can do (Cimpian&Markman 2009; Cimpian&Cadena 2010). In this regard, Cibra and Shamsudheen (2015, p. 701) highlight that:

> «Ostensive naming of a novel object […] provides two types of information to the addressee: a property (i.e., the name) of the kind that the object exemplifies and a label by which other objects of the same kind can be identified».

This entails, according to them, that «an ostensively communicated name will be interpreted as a kind label rather than the name of the particular object» (Ibidem). Furthermore, Csibra and Shamsudheen (2015, p. 698) provide an alternative explanation to Butler and Markman's (2012) results. According to Csibra and Shamsudheen, children's perseveration on the inert blickets is not due to a stronger inference about

generalizability triggered by pedagogical stance, rather because infants have learnt (thanks to demonstration) that blickets, as a kind, are magnetic, and thus, they infer that the inert ones do not work because they are broken. Thereby, Csibra and Shamsudheen (2015, p. 698) are assuming that children interpret the communicative demonstration «as being about a property of the kind exemplified by the blicket used in the demonstration». Through the pedagogical demonstration, children learn something about the property of the kind, and such nonverbally acquired knowledge is resistant to negative evidence, just like Sarah-Jane Leslie and colleagues (2011) have shown about linguistic generic expressions «which are not invalidated by counterexamples» (Csibra and Shamsudheen 2015, p. 698).

Another aspect is worth pointing out: what the children learn pedagogically does not regard only the kind and its main property (e.g., "blickets pick up paperclips"), but also how the kind of object is used by the persons, i.e. by the members of a community. In essence, children infer also about something that has a *social valence* (e.g., "one uses blickets to pick up paperclips"): based on how one manifests ostensively to use novel objects, or in other terms, biased by the pedagogical demonstration children infer how they (and the other people) *should* interact with the novel objects (Butler&Markman 2012, p. 1425-1426). This interpretation leads us to introduce another assumption predicted by the natural pedagogy theory, the assumption of universality.

2.4 Assumption of universality

The *assumption of universality* (termed also *shared knowledge assumption*, or simply *omniscience assumption*) is a bias according to which whatever the child learns (at least) pedagogically is supposed to be known by everyone. Such corollary of omniscience allows that the knowledge that the child acquires is *believed* as public, shared, and universal. «If someone knows something, everyone knows it» (Csibra&Gergely 2006, p. 272). This is the maxim that synthesised the bias that is crucial within the natural pedagogy theory for two reasons at least. First of all, the omniscience assumption is strategic from an evolutionary point of view, because it allows a fast transmission and a strong persistence of social norms, in virtue of the generation of behavioural expectations by the infants about social habits manifested by

others despite their cognitive opacity. It is also connected in the same way to a fast learning and easier sharing of the first words, as suggested by Csibra and Gergely, who claimed in this regard that «a child can plausibly assume that a word learned from a certain person is not her specific way to express a certain concept, but part of a shared sign system» (Csibra&Gergely 2006, p. 272). The second reason strictly linked to the last claim relies on the fact that, according to the assumption of universality, the knowledge content assimilated pedagogically is ascribed to others without any commitment with a (simulationist) mentalisation. The pedagogical learning deals with an object-centred perspective that ignores (or fails) the teacher's mental state. Thus, it would be exactly the opposite of a person-centred perspective predicted by the simulationist mindreading account. Indeed, such assumption implies that an individual behaviour manifested pedagogically as a personal dispositional property is learned by infants and children as an epistemic status (or a knowledge content) that goes beyond the episodic situation in which it has been manifested to be relevantly ascribed to other members of one's social group. In other words, as Gergely sums up: «infants expect ostensively manifested referential information to represent common cultural knowledge that is shared by and accessible to other individuals as well and not only to the communicating person demonstrating it» (Gergely 2013, p. 147).

I suggest that the bias predicted by such assumption is permitted and guaranteed by an early mindreading system that is not based on a simulationist account. I will analyse the features of such primary form of mindreading in Chapter 4 and in Chapter 5. It will be illustrated along the first paragraphs of the fourth chapter. The task I prefix here is to show, through the description of two experiments, the characteristics of this kind of inference triggered by the pedagogical teaching/learning modality and manifested in terms of expectations and comprehension about others' preference-behaviour.

The experiments I am going to illustrate start from the assumption that emotion disposition towards an objects-preference expressed by an individual is treated as an object-directed behavioural manifestation exactly like verbal labelling, and demonstrations of the object-functional properties. Therefore, expressions of emotion disposition can be interpreted by infants as salient information potentially referring both to the object preferred (or rejected) and the subject who makes the preference. Testing

10- and 12-month-olds' behaviour facing to televised positive and negative emotional reactions of an actress, Mumme and Fernald (2003, p. 221) noticed that 12-month-olds (at least) «used social information presented on television and associated emotional signals with the intended target». The authors recognise that it is methodologically different putting the infants in the role of an onlooker to a sequence of events, or, on the contrary, when infants participate in the social interaction. Mumme and Fernald (2003, p. 221) highlighted that «when the child is a participant, the question of interest is how attentional, emotional, and pragmatic cues from the social partner influence the child's own behaviour»[14]. This is what Gergely and colleagues attempted to investigate in order to verify the universality implication of natural pedagogy.

2.4.1 Gergely and colleagues' experiment (2007)

Gergely, Egyed, Király (2007) tested 14-month-old infants with a violation-of-expectation looking time paradigm in order to investigate whether object-directed emotion manifestations - performed by adults through communicative interactions - did modulate infants' own object-directed actions without relying on person-specific mental state attributions. In other terms, given particular conditions and ostensive teaching context, infants may succeed in social referencing «without necessarily relying on or even having to infer and attribute the other's mind a person-specific mental attitude toward the referent» (Gergely 2007a, p. 190). In particular,

> «It is hypothesized that the presence of ostensive cues biases infants toward interpreting others' object-directed emotion manifestations in an object-centred manner, as conveying relevant new information about the referent kind (e.g. that "broccoli is good!"), rather than as conveying *person-specific information* about the subjective mental attitude that the other person holds toward the referent (e.g. "Allison likes broccoli"), as predicted by the simulationist mindreading account» (Gergely 2007a, p. 190).

[14] In their experiment, Mumme and Fernald merged the two roles: as an onlookers, the infants did watch a televised scenario where an actress reacted with neutral, negative or positive emotions to two novel objects; as a participant, infants had the possibility to interact with two real identical objects to those ones seen on television.

In the experiment twelve familiarisation events were presented, and in each one two demonstrators appeared ostensively greeting the infants (exactly sixty-four 14-month-olds) with smiles, eye contact, and other communicative cues. Two target objects were placed on the left (object A) and right (object B) sides of a table in front of a demonstrator, who manifested a positive or a negative emotion towards each of them[15]. A second demonstrator performed the same actions, but the emotion displayed toward A and B were inverse: if the first demonstrator expressed "interest/joy" toward A, the second one expressed "disgust". During the familiarisation trials, the two demonstrators expressed the same kind of emotions towards the two objects. At this point, two different experimental conditions were created: the «asymmetric presentation» condition and the «symmetric presentation» condition. In the former, one of the agents, called «Frequent Person» (FP), appeared nine times, three times more than the other demonstrator called «Infrequent Person» (IP), who appeared only three times. In the symmetric presentation, the two demonstrators appeared six times each one across familiarisation trials. According to the simulationist mindreading account, infants should attribute different person-specific mental attitudes (liking or disliking) to the two demonstrators toward the two referent objects both in asymmetric and symmetric conditions. On the contrary, according to the universality assumption, the more salient information transmitted by FP's manifestations in asymmetric condition should mask the real preference of IP, leading the infants to ascribe FP's preference also to IP. This should be due to the fact that the object A more frequently manifested by FP to have positive valence is represented by infants as *good* (or as *better*) than the object B demonstrated by positive valence three times less often by IP. Therefore, natural pedagogy theory predicts that ostensive cues trigger an «object-centred interpretation of others' referential emotion displays» (Gergely et al. 2007, p. 141) such as if infants learn that the object A is really good, it must be good for everyone. On the base of their informative achievement gained by the more salient FP manifestations, infants ground their expectations both on the object class that the referent belongs to, and on others' behaviour about that object kind. In a nutshell, infants expect that all others' object-

[15] The objects were novel to the infants, who initially showed no differential preference for either of them. For a brief summary of the experiment see Gergely (2007a, pp. 191-192).

directed actions will be similarly driven by the object's valence quality. In Gergely's terms:

> «following the universality assumption of the pedagogical stance infants would show a generalized *object valence based expectation that all people, including both demonstrators*, would choose that object which the infants had come to represent by the end of the familiarization trials as *good* or as *better* than the other» (Gergely 2007a, p. 192).

To verify this prediction, the test trials consisted of four «object-choice and object-directed action» in which the two demonstrators appeared twice but with a neutral facial expression, performing one «attitude-consistent' object-choice», i.e., they chose the object targeted by positive (liking) emotions showed during familiarisation, and one «attitude inconsistent object-choice» in which they chose the object targeted by negative emotion). The chosen object was grasped by the demonstrators that moved it to a new position (just ten cm away), and then they moved it back to the original position. This action was repeated many times until the young subject watched it. According to simulationist account, infants should be more surprised, and thus, they should look for longer at the unexpected choices, namely «the attitude inconsistent object-choices for both conditions» (Gergely et al. 2007, p. 141). In contrast, the results contradict such predictions, because, on the one hand, a sharp looking-time distinction occur between the two conditions (symmetric and asymmetric), and on the other hand, a significant distinction between consistent and inconsistent attitude lacks in symmetric condition. Furthermore, in the asymmetric presentation a longer looking time was recorded only when the second demonstrator chose «the more negative-valenced Object B (M=19.15 sec)», that was consistent with her original choice showed in familiarisation trial, but inconsistent with «the more positive-valenced Object A (M=15.50 sec)» made by the first demonstrator (Gergely at al. 2007, p. 144). In particular, the final result recorded in the asymmetric condition (indicated with the term «Object-Valence main effect») may represent the strongest support to the universality implication of natural pedagogy theory. It should be interpreted as the clearest evidence that infants did not attribute person-specific mental attitudes to the two demonstrators or that although 14-month-olds are allegedly able to *read* mental states, they seemed to not use the person-specific

mentalistic information to predict the demonstrators' actions during the test trials (Gergely 2007a). Indeed, Gergely and his co-operators don't claim that infants (at least at 14 months of age) are unable to ascribe mental states to others. Rather, they proposed (see also Csibra&Gergely 2006) that infants are directed by the presence of ostensive cues that triggered an *object-centred interpretation switch*, or in other terms, ostensive cues constrain and drive infants' interpretation of the others' object-directed communicative actions such as, for instance, object-referential emotion manifestations, verbal and non-verbal explanations of functional object properties, verbal labelling. Therefore, on the basis of this experiment, the authors underlined the crucial role played by the ostensive cuing context as the main factor enable to bias infants to interpret the manifested emotions as transmitting information about the referent qualities, and not about the other's subjective mental attitude towards them.

However, Gergely and colleagues' study (2007) remains inconclusive insofar as they tested the contrasting interpretations of emotion's expressions only in the context of a communicative manifestation. Indeed, we could wonder whether in absence of the binding ostensive cuing context the infants are able to mentalise and, thus, understand object-directed emotion expression as subjective emotional disposition towards the objects (as previous approaches tried to show, see e.g. Moses et al. 2001; Mumme&Fernald 2003). Egyed and colleagues (2013) attempted to answer this question testing 18-month-olds in particular experimental conditions with the aim to figure out «the special power of ostensive signals to induce a non-episodic interpretation of a communicative agent's object-directed emotion gestures as conveying relevant information about motivational dispositional properties such as preferences that are socially shared» (Jacob&Gergely 2012, p. 70).

2.4.2 *Egyed and colleagues' experiment about preferences socially shared*

The study conducted by Egyed, Király and Gergely (2013) is similar to the prior study just explained above, but it has the merit of comparing more strongly ostensive communicative condition with non-communicative demonstration context. Indeed, by presenting object-directed emotion manifestations, they demonstrated that 18-month-old infants are able to flexibly attribute either an object-centred interpretation or a person-

centred interpretation to referential emotion displays respectively in communicative and non-communicative contexts. In their experiment, 18-month-olds look at an adult manifesting positive and negative object-directed emotional expression (interest vs. disgust)[16] towards two novel objects of different colours and shapes placed on a table, one on the left side, and the other on the right side. Three groups of infants have been created, one for the ostensive communicative condition, and two groups for the non-communicative experimental condition. In the ostensive communicative condition, first of all, a female demonstrator addressed each infant (seated in mother's lap in front of the table) by eye contact, smiles, calling him or her by their name in infant direct speech. At this point, the agent, turning to look at one of the unfamiliar objects, displayed a positive facial-vocal emotion expression (joy/interest), and then manifested a negative emotion expression (dislike/disgust) towards the other object. This sequence of events was repeated twice. On the contrary, in the non-communicative condition, the demonstrator neither looked at the infant nor addressed her/him by IDS, performing twice the same sequence of acts as if she were alone. In the test phase, infants of the group in ostensive communicative condition, and those of one group in non-communicative condition saw another female agent (the *requester*), who was not the same person of the demonstrator, who came in without paying attention to the objects, and immediately requested the infant to give her one of them while she was looking at him/her eyes. Only in the third group of infants, the requester was the same person seen during the familiarisation trial in non-communicative condition. The results revealed «a significant difference in the distribution of objects' choices among the three conditions» (Egyed et al. 2013, p. 1351). The most significant difference in infant object choice (made by touching or grasping) was recorded between non-communicative/different person condition (in which «5 infants chose the positively valenced object, and 11 infants chose the negatively valenced object») and non-communicative/same person condition (in which «14 infants chose the positively valenced object, and 2 infants

[16] Following Repacholi&Gopnik (1997) and Repacholi's (1998) procedure for identifying the target of an emotional display. In the latter study, infants were presented with two boxes, and each box contained an object that could be visible by opening the box lid. «An experimenter expressed happiness as she looked or put her hand inside one box, and disgust as she repeated this action with the other box. Infants touched both boxes but preferred to search for the happy object» (Repacholi 1998, p. 1017).

chose the negatively valenced object» (Egyed et al. 2013, p. 1351)). These results may represent a further demonstration that by observing someone's object-directed emotion, infants (at least at 18 months of age and presumably before) are able to ascribe to that individual the corresponding emotion accordingly to the person-specific emotional attitude state.

By the comparison of the object choices in the communicative-context/different-person condition («11 infants chose the positively valenced object, and 5 infants chose the negatively valenced object»), the researchers found a significant difference with those choices accomplished by infants in the non-communicative-context/different-person condition (Egyed et al. 2013, pp. 1351-1352). In conclusion, Egyed and co-operators found that when infants observed the referential emotion's expression from a third-person perspective, i.e. in a non-communicative setting, they assigned a person-centred interpretation. Such assignment avoided applying the agent-specific attributions to other subjects. Infants did not appear to generalise the object choice. On the contrary, when infants were involved in a second-person perspective, they assigned an object-centred interpretation to emotion displays. In virtue of communicated ostensive signals, «infants readily generalized their interpretation of the communicative agent's referential emotion manifestations as applicable to other individuals as well» (Egyed et al. 2013, p. 1352). In this way, infants may represent the information achieved as shared knowledge available to others, «forming part of the cultural common ground shared by one's social community» (Egyed et al. 2013, p. 1349). On the one hand, universality assumption allows infants to modulate their future behaviour in an emotion-congruent manner when they will encounter referents belonging to the same kind. On the other hand, on the basis of universality assumption, infants expect that also who have not manifested locally the emotional attitudes towards the referents have the same disposition and judgment toward the same kind of referents.

2.5 Universality without mindreading?

The advocates of the theory of natural pedagogy do not claim that infants develop later mindreading skills, or that they are unable around the first year of age to ascribe epistemic mental state, emotions, intentions, desires, or dispositional attitudes.

Rather, they affirm an *independence* between pedagogical processing and mentalizing skills to infer and represent mental states of others. I report Gergely's thought: «*The central claim* that pedagogy theory makes *is independent* [from] the question of whether and/or at what point in development young infants are able to infer and represent mental states of others» [my italics] (Gergely 2007a, p. 193). The conclusion prefigured by Gergely is that «one can learn from other minds without learning about them» (Ibidem).

> «The basic proposal of natural pedagogy is that young infants possess a dedicated cue-driven social learning system that very early on enables them to fast-learn relevant - even if cognitively opaque - cultural knowledge from other minds without necessarily attributing mental states to those minds. *This certainly does not preclude the possibility that young infants may already be able to infer and reason about others' mental states*» [my italics] (Gergely 2007a, p. 193).

It has been sustained by Gergely, Csibra (Csibra&Gergely 2006) and colleagues (Gergely et al. 2007) that the development, by establishing the comprehension of other minds, involves the suspension and the inhibition of the default universality assumption whereby other minds are seen as equally omniscient. Such gradual learning process driven by experiences should lead to the mature mentalising ability allowing to represent the differential knowledge contents of individual minds. I suggest that it is true only in part. In fact, if we embrace a model of infant's mindreading that enables to serve the cognitive skills necessary to support omniscient implication, and if we imagine that this core innate mindreading changes and develops its functions during the childhood reaching the maturity step by step and driven both by experiences and by executive functions onset, then we can define a dynamic and jarring cooperation between mindreading and natural pedagogy systems. Much more needs to be done in order to analyse the evolution and the supposed inhibition of natural pedagogy system in favour of theory-of-mind skills onset enabling infants «to appreciate that the people around them possess separate minds with differential knowledge contents that represent the world in different ways» (Gergely 2007a, p. 172). Further findings should investigate also at which developmental stage the decrease of deferential attitude occurs (see § 2.2 and Chapter 6). In fact, differentiating others' minds and modifying the children sensitivity towards ostensive cues should entail the inhibition or the suspension

of the epistemic trust in respect of the knowledge source that is not seen as blindly reliable anymore. This is a crucial developmental acquisition that (should) make children more vigilant towards those unreliable, uninformed, deceptive communicators subjects (Gergely et al. 2007), but its downward trend has to be tested more clearly in a pedagogical context.

In the next chapter, I will examine in more details the strength and flexibility of early belief attribution also in non-pedagogical context in order to sketch a potential model of early mindreading system able to serve universality assumption, but for the moment it is better to see if other findings may provide major support and confirmations along child development, on the one hand, and in the other hand, if pedagogy respects its innate and human-specific characteristics among other cultures, and among other non-human animals from an evolutionary perspective.

3. Investigating natural pedagogy in pre-schoolers

In order to learn much more about natural pedagogy system, it is indispensable, in my opinion, to analyse its onset and the dynamic ongoing of its implications (from the epistemic trust to universality assumption) as biasing factors for cognitive functions along the first years of ontogenetic development. I have already recalled above, in § 2.3.2, through Butler and Markman's experiment, natural pedagogy in preschoolers. Sobel and Sommerville (2009) noticed that a pedagogical context elicited 4-year-old children to learn more successfully a sequence of lights rather than children provided with an inappropriate rationale or no rationale ordering style. Bonawitz et al. (2009) further reported «how pedagogy helped preschoolers focus on a specific function of a toy over-and-above accidental exposure» (Sage&Baldwin 2012, p. 154, 156). Sage and Baldwin's studies have improved natural pedagogy's theory from USA (and beyond the Gergely and Csibra's staff and usual co-operators), enhancing above all the role of pedagogical cues for providing «unique assistance to infants in terms of boosting their production of causally effective action with [tools]» (Sage&Baldwin 2011, p. 836).

3.1 Sage and Baldwin's experiment

Sage and Baldwin (2012) examined natural pedagogy's effects on thirty-two children of different ages (3- and 4-year-old) involved in a play situation with their parents into a natural pedagogy environment. The experimenters constructed two toys (Pyramid and Flops) that were completely novel for the parents and the children who were engaged in four toys tasks that included: «(1) children being taught the functions of either the Pyramid or Flops toy by a parent, (2) children teaching the functions of that toy to the experimenter, (3) children being taught the functions of the other toy by the experimenter, and (4) children teaching the functions of that second novel toy to their parent» (Sage&Baldwin 2012, p. 160). Parents were divided in two groups, half of parents were assigned to teaching activities, and to this aim they were informed by the experimenters about the functions of one novel toy (or Pyramid or Flops). The other parents half were informed that they would have been involved with play activities. Before trials children and parents played free together for five minutes in a room where there were fifteen toys on the floor of a diverse nature like a train, a puzzle, bus, keyboard, jack-in-the-box, bouncy ball, noise-making ball, Furreal puppy. Then, session trials started and only one adult at a time entered the room with the toy they would be teaching. Only after tasks 1 and 3, children were instructed by the experimenter that they had the job to teach the other parent about the toy that he/she had never seen before. But whereas the experimenter taught all the four toy's functions, the parents did not necessarily demonstrated everything.

This intriguing experiment provide a lot of suggestions and insights. First of all the authors sketched *how* pedagogical parents appeared to be during both the free play session and the novel toys session making a score of use and frequency of the several cues adopted. Sage and Baldwin's analyses indicated that pedagogical cue use was «omnipresent across contexts», and parents «seem to capitalize upon multiple pedagogical cues» (162-163): eye contact, gaze shifting, joint attention, name referral, pointing, referential speech (e.g. "Look!" or "Watch!"), suggestions (e.g. "Do you want to put the puzzle together?"), knowledge questions (e.g. "What does it do?"), observations (e.g. "Oh, you like that one"). The authors found a predominant use of speech and a strong interrelation between pedagogical cues with the exception of eye contact and name referral above all. At first glance it may be quite surprising given the

fundamental role played by these two cues in natural pedagogy, but the age of children's sample may explain such result, suggesting the idea that – as Sage and Baldwin pointed out - «pedagogy might be best viewed as a *multidimensional construct*, with differing cues subserving distinguishable subgoals [my italics]» (Sage&Baldwin 2012, p. 173). As predicted by Csibra and Gergely's theory the pedagogical context in play sessions had meaningful effects on children's learning, since 68 % of children learned all four functions when taught by a parent. On the contrary, in a separate group of sixteen children left to their own exploration without a teacher, only 19% learned all functions. It is worth noting that if, on the one hand, pedagogy favours children to broaden their knowledge, on the other hand, pedagogy constraints their free explorations as pointed out also by Bonawitz and colleagues (2011), whose experiment documents that preschoolers may be very sensitive to pedagogy in play contexts, but at the same time it shows that children played less with a novel toy after witnessing a demonstration of a single function in a pedagogical way.

Sage and Baldwin study analyses also the relations between parents' and children's use of pedagogical cues for teaching each other toys' functions. During the child-taught toy task the researchers tried to find correlations with the relative parent's pedagogical approach. Their analyses revealed «that parents' use of referential speech was positively correlated to children's use of referential speech in the child-taught toy task» both with the parent and with the experimenter, but there were no other significant correlations (Sage&Baldwin 2012, pp. 171-172). Therefore, children's use of pedagogical cues during the child-teaching tasks was unrelated to their parent's use of those cues during the parent-teaching task for both toys, flops and pyramids. This result deserves further researches, anyway it may suggest that child pedagogy has its peculiarities regardless of adult teaching modalities.

4. Natural pedagogy in a teaching evolutionary perspective

Natural pedagogy is part of a long tradition of developmental studies investigating the shapes and the dynamics of child-parental communications aimed to social learning (Bloom et al. 1976; Bruner 1977; Iverson et al. 1997), and it extends the recent experimental literature documenting infants' ability at utilising adults' communicative

signals to guide their inferential processing in learning contexts (see for example Baldwin 1991, 1993; Carpenter et al. 1998; Moses et al. 2001). Beside such cognitive perspective discussed so far, natural pedagogy arises questions about teaching from an evolutionary point of view. This issue is treated by Csibra and Gergely in their first papers (Csibra&Gergely 2006; Gergely&Csibra 2005; Csibra 2007b), and I would like to focus first of all on the adaptive value of teaching.

Teaching solves several adaptive problems arising in social learning that cannot be addressed by learners' behaviour alone. As we have seen, in order to learn, pupils must attend to and have access to novel information that is normally available through experience, observation or direct communication. The latter facilitates naïve learners to catch faster the relevance of information transmitted by the teacher. There is a broader sense to intend teaching comparable with nonhuman animals, as Caro and Hauser (1992) tried to draw in their seminal study providing three criteria for defining teaching in nature:

1) an adult, or an expert individual changes his/her behaviour in the presence of a naïve conspecific observer.

2) The expert individual spends efforts and time without immediate benefits.

3) However, in virtue of his/her behaviour, the naïve observer acquires competences and knowledge rapidly and more efficiently.

Such definitions, fully applying to human natural pedagogy, makes teaching different from other forms of social learning processes, above all under the respect of a necessary cooperative interaction (Thornton&Raihani 2008, p. 1825; Galef&Laland 2005). A cooperative interaction entails communication, and we may claim surely that *teaching is always a form of (intentional) communication*, but it is different from other communicative forms, insofar the teachers' influence affects pupils' behaviour. According to the first criterion, teaching entails a behaviour projected to aid and promote learning in another. This is properly defined as mentalistic approach to teaching activity (Klein 2015), whereby it could appear that teachers need mindreading system to identify the very need for teaching, «to figure out what it is that they ought to teach, and to tailor the difficulty of the task to match the skill level of the pupil» (Klein 2015, p. 2; Kruger&Tomasello 1996; Tomasello et al. 1993). Ziv and Frye (2004) sustain that teaching is *only* possible when teachers and learners consciously recognise

both intentionality and knowledge differences between individuals. Olson and Bruner (1996) attribute to theory of mind the ascription of ignorance that represents the starting point of attempting to teach; according to Strauss et al. (2002, p. 1476): «in order to teach, one needs to know when knowledge, beliefs, skills, etc. are missing, incomplete, or distorted, as well as how people learn».

Against this strong anthropocentric perspective it is worth pointing out that the three teaching criteria mentioned above are checkable and sometimes evident in some species like, e.g., ants, bees, meerkats, domestic fowls, domestic cats, pied babblers (Hoppitt et al. 2008; Caro&Hauser 1992), but not among great apes (Caro&Hauser 1992; for a brief review see Thornton&Raihani 2008; for an opposite and controversial interpretation about teaching in nonhuman primates see Boesch&Tomasello 1998). Here, I am not claiming that among great apes doesn't occur social learning and cultural transmission, where "cultural" means different social habits in different groups. Indeed, there is agreement among scientists about the fact that chimpanzees learn from one another, but it is still hotly debated whether it is the case to speak properly about teaching according to the three criteria (for a review see Moore 2013; Scheel et al. 2015). So far, empirical evidence supports the hypothesis that «the primary mechanisms of chimpanzee social learning are […] combinations of affordance learning, local and stimulus enhancement, and emulation» (Moore 2013, p. 889).

The element that distinguishes human teaching may be more parsimonious than that one advanced by mentalistic teaching theorists and it regards its great flexibility suitable in many different contexts (Thornton&Raihani 2008, p. 1828). Human teaching promotes both procedural (or normative knowledge referring to knowing-how), and declarative (or descriptive knowledge referring to knowing-that). In the case of natural pedagogy, we may notice an overlap of these two dimensions of knowability[17]. However, the presence of such twofold teaching aspects, so strongly interconnected in human cultural transmission, has been noticed also in some ants (Temnothorax albipennis; see Franks&Richardson 2006) and birds (pied babblers, see Rahini&Ridley 2008), and also in vervet monkeys whose use of positive reinforcement is supposed to

[17] See Vorms (2012, p. 539) who turns into a problem such overlap, which has not to be assumed, but rather it has to be better analysed, above all when infants deal with certain kinds of tasks (like A-not-B task) where the actions involved by the agent can be easily misunderstood (see Chapt. 6).

facilitate pupils' learning about alarm calls (Caro&Hauser 1992). This evidence suggests the usefulness of comparing animal and human teaching in order to provide a broader notion of teaching, and suggests abandoning the anthropocentric point of view to better see the peculiarities and the continuity of human beings transfer information through evolution. This evidence represents also a further element about individual cognitive foundations of teaching (Thornton&McAuliffe 2012), and they may sustain, in my opinion, the suggestion that none inference on teacher's mental states appears to be necessary to spark teaching/learning relationship and also to trigger pedagogical stance, at least under its *primary relational features* (based on the link between teacher's effort and learner's attitude), and the *general knowability categories* (normative and semantic levels). Indeed, if teaching is a form of communication that shall strictly respect the Caro and Hauser's three criteria, natural pedagogy may be seen as the human variant reliant upon specific human communicative cues, whose comprehension is not necessarily tied to, or triggered by, mindreading capacities.

Putting natural pedagogy within a functionalist evolutionary non-anthropocentric notion of teaching brings to rethink the general assumption of the theory whereby the motivation to teach is specifically human (as otherwise claimed by Csibra&Gergely 2006; 2009, but see also Strauss&Ziv 2012)[18]. However, I would not underestimate the unique character of human pedagogical teaching allowing the transmission of cultural and *generalizable* knowledge with opaque content (Csibra 2007b), as pointed out also by zoologists (Thornton et al. 2007). Indeed, the species mentioned above perform types of non-human teaching that «are unable to support the transmission and maintenance of cultural forms with *unrestricted content* because the content of transmitted information is either strictly pre-specified (e.g. the location of particular food source) or left to be discovered by the pupil [my italics]» (Csibra 2007b, p. 96). Furthermore, following the taxonomy provided by Klein (2015), if natural pedagogy can be defined without any doubt as a "direct active teaching" (DAT) akin to other teaching modalities observed among other non-human animals, at the same time, natural pedagogy is human-specific among social learning practices because the teacher does not merely play the role of attractor of pupil's attention towards certain actions, or objects. Rather, through ostensive cues, the teacher «manifests to the pupil that she is the intended addressee of

[18] For a functionalist teaching account see Klein (2015).

the demonstration» (Tanone&Csibra 2015, p. 50). Acknowledging the existence of DAT in non-human animal kingdom and the use of referential signals (e.g., alarm calls, food calls, bee dance, etc.), human ostensive communication is unique because it allows acquiring generic knowledge contents that are opaque material and social kinds, artifacts and conventions, not functionally transparent, and without any evident fitness value. The evolution of such cognitive adaptations allowed pupils to interpret the taught information as being applicable *beyond its episodic use and contextual achievement*, whereas the referential signals used in non-human DAT are «restricted to episodic facts in the 'here and now' and cannot express content that is generalizable to other situations, other locations or other individuals» (Csibra&Gergely 2011, p.1150).

> This is precisely the type of inference that ostensive signals license about demonstrated content. Therefore, if human teaching is to be portrayed as a glaring exception in the animal kingdom, this is not, or not solely, because of its frequency and breadth of use, but rather because of its capacity to perpetuate cultural kinds that are causally and teleologically opaque (Tanone&Csibra 2015, p. 50).

5. *Evolutionary origins of natural pedagogy*

Csibra and Gergely (2006; 2009; 2011) provide an explanation of the evolution of teaching in terms of adaptation. However, their approach draws teaching as a uniquely and universally human adaptation. According to them, natural pedagogy represents «an evolutionary adaptation along the hominin lineage» (Csibra&Gergely 2011, p. 1149).

Human instrumental actions and social conventions include almost always opaque elements. Actually it is quite difficult to understand immediately the use, the efficacy and the adaptive value of eating something by a plate and a spoon hard to handle. Born and immerged in human cultures, infants see and observe constantly so many actions that appear to be (and actually are) arbitrary, even counterproductive, but at the same time, they are relevant and necessary to be learnt to establish fundamental social exchange and promote a normal style of life. Therefore, the intergenerational transfer and the consequent cultural stabilization of human technological habits, social conventions, folk traditions actually pose «a learnability problem for the purely observational learning mechanisms» (Csibra&Gergely 2011, p. 1151). The role of a

benevolent teacher, here intended as a guide who marks and emphasises the relevant aspects of these actions through specific communicative signals, is necessary. On the other side, an innate mind architecture equipped with dedicated cognitive skill is necessary too.

The origins of natural pedagogy – interpretable as testimony processing, in the sense of «transmission of observed information» (Goldman 1999, p. 103) - must be connected to human communication modality, and maybe in particular to the need of manipulation of others' mental state (Sperber 2001). Perhaps more simply, as suggested by Tomasello (2008, p. 6), communication is fundamentally a «cooperative enterprise», i.e., it operates «most naturally and smoothly within the context of […] mutually assumed cooperative communicative motives». Therefore, we may suppose that at a certain point of human evolution the cooperation among humans became more extended eliciting a strong evolutionary pressure for the emergence of specific forms of communication in which the transfer of knowledge was necessary. However, this is not enough to explain how natural pedagogy system arose, for instance, with its specific characteristic skills to generalise information learned. According to Csibra and Gergely (2006; 2011), the determinant factor for the evolution of pedagogical relation was the production and employment of more and more sophisticated tools. It is not impossible but quite difficult to understand and learn properly by passive observation or trial-and-error those instrumental actions and artefacts that «tend to be opaque both in terms of their adaptive function (*teleological opacity*) and in terms of their modus operandi (*causal opacity*)» (Csibra&Gergely 2011, p. 1154). Direct communicative demonstrations serve the task to guide pupils towards the relevant information and important aspects of the situation, and at the same time the non-verbal and verbal "instructions" favour the acquisition of information as generic knowledge that goes beyond the episodic and contextual artefact use observed in the demonstration. Then, such acquired information is applicable to other circumstances and other objects of the same kind. This presupposes an adaptive cognitive equipment in infants probably matured in such cooperative social and technological environment where a lot of interactions are subserved between adults and offspring, and not only between a mother and her own babies whose care has been shared within a group at least for more than a million years (Hrdy 2009). The emergence of such cooperative breeding system

constitutes the distinctive character of human evolution (Burkart et al. 2009) and it probably represents the main general factor, together with neurobiology and technology, that made possible and necessary the natural pedagogy onset.

6. Natural pedagogy is universal

An important implication of natural pedagogy as the evolutionary result of cognitive adaptation to human social environment is that «it must be universal across human cultures» (Csibra&Gergely 2011, p. 1152). Anthropologists like Lancy (see e.g. Lancy et al. 2009) are skeptical about this claim, nevertheless some of their reports about case studies of teaching modalities among children belonging to non-Western societies (Kpelle community in Liberia) seem to confirm Csibra and Gergely's hypothesis. Rather, the anthropological researches (see also Hewlett et al. 2011) point out that it is the frequency of teaching that significantly differ between Western and non-Western societies and the reports descripted reveal how many of the predictions of the theory of natural pedagogy are confirmed in several traditional cultures (Csibra&Gergely 2011, p.1153). Furthermore, compared studies on *overimitation*[19] among children belonging to non-Western society (Nielsen&Tomaselli 2010) show how they display similar learning biases to Western children when they deal with novel actions in communicative context.

7 Natural pedagogy and psychological essentialism

The assumption of generalizability predicted by natural pedagogy fosters essentialist view insofar as children are fast and efficiently guided to categorise an object as a member of a kind. This elicits the inference (i.e., the induction) that invisible and influential properties of an entity are shared by and extended to the members of the category. Psychological essentialism consists in the belief that entities have essential

[19] *Overimitation* consists in the reproduction, after infants' observation of an adult operating a novel unfamiliar object, not only of those acts causally necessary but also those clearly superfluous to the end-aim (Lyons et al. 2011). Overimitation often occurs in pedagogical conditions, when infant follow faithfully the demonstrator's action rather than in other social learning conditions.

causal properties which are hidden and not immediately perceivable, but they are responsible for the observable features of the entities. Essentialism is described as a human «reasoning heuristic» (Gelman 2004, p. 404), or bias (Gelman 2003) that deeply affects conceptual categorisation (Gelman 2003) across cultures and contexts. As Medin and Ortony (1989) suggested, the hidden and causal essence is attributed to the categories to which entities belong. Therefore, following Gelman (2004, p. 404), essentialism can be defined as the view according to which «certain categories have an underlying reality or true nature that one cannot observe directly but that gives an object its identity, and is responsible for the similarities that category members share».

In the initial stages of development, it seems to infants that the world is divided into a few global broad classes of things, such as, for instance, animals or vehicles: «Animals are things that move themselves and act on other things; vehicles are things that give rides» (Mandler 2003, p. 115). This entails that:

> «infants generalize from familiar instances such as dogs to the entire domain of animals, and from cars to the entire domain of vehicles. This result means that physical similarity between the observed exemplar and the generalized exemplar did not play an important role; infants generalized from dogs to cats, rabbits, fish, birds, and aardvarks (but not to vehicles) and from cars to trucks, buses, motorcycles, airplanes, and forklifts (but not to animals)» (*Ibid.*, p. 109).

Therefore, at the beginning of their exploration and encounters of the world, infants tend to generalise in as such broad way on the basis of abstract, undifferentiated conceptualisations of vehicles, plants, animals. Only with experience and with the ongoing language acquisition, infants learn to differentiate entities in finer categories, and to notice perceptual details making more accurate inductive inferences (*Ibid*., pp. 114-115).

7.1 Natural-born essentialist

As Paul Bloom (2010, *Preface*) said: «We are natural-born essentialists». There is an essentialistic compulsion to consider the world as constituted by immutable forces and facts; such psychological attitude permeates common sense in every experience

domain (Perconti 2015, pp. 101, 110). In this sense, essentialism can be depicted as an evolutionary, adaptive shortcut for facing and handling, classifying and using the surrounding entities and events. As Gelman (2004) pointed out, the potential induction of categorisation generates expectations about entities' behaviour.

The two biases, triggered by natural pedagogy attitude, promote and encourage the essentialistic compulsion that affects so profoundly the common sense. In fact, attributing to others generalised aspects of object or social categories, for instance, can strengthen the essentialist bias in a particular way when the expectations about others' behaviour are confirmed. Firstly, children generalise pedagogically discovered features or functions of an object and extend them to its category. Secondly, children assume as true for everyone the generic acquired knowledge, so they expect that people behave coherently with such knowledge. The combination of the two pedagogical biases, above all when they are accompanied by ordinary and empirical confirms, reinforce certainties that common sense needs. Those certainties that: *i*) have the form of beliefs (propositional representations with simple content), *ii*) belong to cultural and ordinary knowledge about the surrounding world, and *iii*) are considered common among the members of community *in force of the fact* that such beliefs are assumed to be shared by other people. I am not proposing an epistemological (and specifically causal) hierarchy between essentialism and natural pedagogy assumptions. Rather, I suggest that essentialist bias cooperates with the generalisation and universality biases mutually reinforcing their inferential power.

Furthermore, in the domain of folk psychology, the psychological essentialism take the shape of the '*attributive dualism*' according to which we normally represent animate entities as they were guided by mentalistic reasons (Perconti; 2017, pp. 46-48; 2011). In this regard, the universality assumption match, and (may) sustain at the same time, such representational disposition to ascribe mental states to others. The spontaneous ascription of beliefs, that for me grounds the universality assumption generating behavioural expectations, may work harmoniously with attributive dualism.

They corroborate each other. In respect of such innatist aspect, we can claim, following again Paul Bloom, that we are «natural born dualist» (Bloom 2004)[20].

The natural inclination (so spread among children and adults) to represent animate bodies as mentalistic creatures elicits dramatically a flexible ascription of propositional contents. In this way, omniscience assumption may contribute to establish and foster some social trivial prejudices. For example, the strong cultural belief that (all of) girls prefer the colour pink. On this base, I entitle to reason as follows: I expect that girls love wearing pink dresses, and it would be very strange and surely unexpected to see a guy completely pink-dressed! *That* boy *betrays* his belonging, i.e., the category who has to belong to, as a guy himself.

Susan Gelman (2003) rejects the suggestion that children conceptualise all concepts. In her view, naïve essentialism is limited to natural (including water or gold, animals and plants) and social kinds (e.g., gender and race); furthermore, its origins emerge late in developmental cognition, even during school age. However, natural pedagogy entails that children project essences also to artifacts kinds (including every made-man objects). The suggestion advanced by Futó and colleagues (2010) (i.e., preverbal infants are able to generalise information about artifact kinds achieved from instrumental non-verbal actions performed by communicative agents) goes against the Gelman's account. According to Jacob and Gergely (2012, pp. 80-82), not only evidence provided by Futó et al. (2010), but also the results of Butler and Markman's (2012) findings, Gergely et al.' experiment (2002) and Topál et al. (2008; 2009) researches show that nonverbal demonstrations, if well suited through the proper use of communicative acts, convey generic information about artifact kinds that might be classified in a essentialistic manner. The essentialist construal of artifact and social kinds would be permitted by the very natural pedagogical procedure: firstly by human infants' tendency to epistemically trust their benevolent communicative informants/teachers, and then by the affection of the two further pedagogical biases. The epistemic opacity of artifacts, which populate the social ordinary world, «may encourage infants to assume, in accordance with the essentialist bias, that the surface-

[20] Cit. in Perconti 2017, p. 56. I thank prof. Pietro Perconti for the suggestion to connect both natural pedagogy biases to psychological essentialism.

observable properties of man-made tools result from their underlying essential properties (e.g., their intended function)» (Jacob&Gergely 2012, p. 83).

8 Conclusions

Natural pedagogy is more sophisticated than other kinds of social learning like, for instance, observational learning in which infants simply respond with increased attention (Sage&Baldwin 2010). Adults through a set of ostensive-pedagogical cues like eye contact, gaze shifting, pointing, IDS, elicit in infants and children the pedagogical learning stance, «a specific attentional and interpretative mindset […] that shapes their processing in ways that expedite learning» (Sage&Baldwin 2012, p. 154). In pedagogical relationships, the naïve learners alter, reshape event-processing eliciting infant's capacity to use the information gleaned from "teacher's" demonstration in their own action and action prediction (Sage&Baldwin 2011, p. 840; Sage&Baldwin 2012, p. 155), according to two functional assumptions concerning the presumption that the pedagogically transmitted knowledge contents are both generalizable information (about referent kinds) and universally shared by other people. According to such *default assumption*, the communicative agent becomes «sources of universally shared cultural knowledge» (Gergely et al. 2007, p. 145). The emerging puzzle regards the fact that if such assumption representing the others as possessors of equivalent omniscient minds, begets a cognitive conflict in learning contexts with one of the proper and main mindreading features: the capacity to distinguish the others' minds as different holders of knowledge contents, intentions, desires, emotions.

The task prefixed for the following chapter is to provide a model of early infant mindreading that doesn't conflict with natural pedagogy, rather it may constitute a *necessary* co-operator to make possible the universality assumption. In fact, what emerges from the analysis of two functional assumptions, epistemic trust and omniscience, leads to the conclusion of an *impossible* cooperation between a full-fledged mindreading system and the natural pedagogy system. The inhibition (or even the suspension) of the two biases implicated by natural pedagogy - the (almost totally) blind trust towards the pedagogical sources of knowledge, and the automatic attribution of the learned cultural knowledge to other people - constitutes an unavoidable

developmental goal for normal subjects. At this point, we could provide two explicative strategies for trying to ground the 'cooperation-hypothesis':

1) the early infant mindreading system is innate and equipped with few skills relative to belief attributions, that change along the first years of life (as advocates of reach nativist account propose, see for a review e.g., Carruthers 2015);

2) or it is the case to introduce a dual mindreading system whereby we would have a primary theory-of-mind system with very different and autonomous features from a later and full structured one (as proposed, for instance, by Apperly&Butterfill 2009). I will argue in defence of the first solution, and against the second one[21].

[21] There is another consideration to do, and it regards the hypothesis that is incorrect to speak about innate mindreading competencies for early infants at all. The advocates of such proposal argue that the early abilities of social cognition (like for example infant false-belief attribution) do not involve metarepresentational skills but enactive capacities acquired by experiences in social interaction, and that the genuine mindreading manifestations (occurring by four years age) are elicited by the development of language (e.g., see Heyes&Frith 2014; for more details see the next chapter).

CHAPTER 4.

A primary form of representational mindreading system.

1. Introduction

Understanding others' mental states means recognising someone's intentions, desires, and beliefs. As Gergely nicely puts it:

> «Young children develop an early comprehension that other people have minds that represent the world. They come to appreciate that others' actions are best understood as being caused by the contents of their representational mind states and the mental attitudes they hold toward these contents. How this inferential and representational mentalizing capacity develops during the early years, and how infants come to appreciate that the people around them possess separate minds with differential knowledge contents that represent the world in different ways, is clearly the second major challenge that an adequate theory of early social-cognitive development must address» (Gergely 2007, p. 172).

In the present dissertation, I refer to such ability indifferently with the term *theory-of-mind* (TOM) or *mindreading*, or *folk psychology*[1], but I have chosen to limit the domain of such competence to *belief attribution*. I will not analyse deeply, for example, desire-reasoning that it is closely connected and, as someone claimed, is «developmentally prior to belief reasoning» (Steglich-Petersen&Michael 2015, p. 526). This choice is motivated by the intention to focus the analysis on the alleged *basic representational structure of mindreading system*. I will attend to Mareschal and

[1] This terminological equivalence rests on a tendency emerged in recent years. As Marchetti and Sangiuliano Intra (2015) noticed, in the literature the concept of theory of mind has frequently been replaced by the concept of mindreading because it reflects more precisely the interpretation of a procedural construct of mentalization pointing to the changes of metarepresentational capacities across the life-span (Marchetti&Sangiuliano Intra 2015). I also embrace such perspective in the present research, so I have opted for the equivalence.

Bremner's advice (2009) to follow the traditional approach that prefers to describe in depth some infant competences.

Following a representational nativist account, I assume that when infants interpret the agent's actions, they usually consider not only his/her motivation, but also *the knowledge that the agent possesses about the scene*. Many researchers have investigated the infants' ability to reason about others' epistemic states.

In the original, and, by now, classic false belief task designed by Baron-Cohen, Frith and Leslie (1985), children listen to the following story of Sally and Ann enacted by the experimenter with props. In front of Sally there are a basket and a box. She hides a marble in the basket and then she leaves. In her absence, Anne moves the object in the near box. At this point, the storyteller asks children where Sally will search for her marble when she returns. The researchers discovered, and their results were confirmed many times later (e.g., Wellman et al. 2001; Rakoczy et al. 2015), that children answer correctly pointing to the basket (where Sally falsely believed the marble is) only starting from the age of four. In contrast, prior to this age, they usually point to the box, i.e., the marble's actual location that they have just heard and seen in the enacted story.

However, since 2005 a great amount of data has been produced by several labs that used and invented different nonverbal methodologies and measures showing that even infants between the ages of 6 and 18 months are able to represent and reason about others' false beliefs (in succession just a few examples arranged in chronological order: Onishi and Baillargeon 2005; Surian et al. 2007; Song et al. 2008; Buttelmann et al. 2009; Scott&Baillargeon 2009; Kovács et al. 2010; Scott et al. 2010; Southgate et al. 2010; Träuble et al. 2010; Luo 2011; Senju et al. 2011; Knudsen and Liszkowski 2012; Scott et al. 2012; Baillargeon et al. 2013; Barrett et al. 2013; Kampis et al. 2013; Buttelmann et al. 2014; Southgate&Vernetti 2014; Buttelmann et al. 2015; Scott et al. 2015; Scott 2017). In other words, also infants younger than 4 seem to hold the same representational capacities although they cannot express in identical manner.

The systematic failure of standard false belief task has brought out a critical issue, namely «whether infants (a) are fundamentally egocentric and as such incapable of attributing to an agent a representation of a scene that differs from their own or (b) are non-egocentric and able to recognize, at least in some situations, that an agent's

knowledge about a scene may be less complete than their own» (Baillargeon et al. 2016, p. 170). Several implicit false-belief experiments seem to indicate that at least by 6 months of age, and certainly around the first birthday, infants are non-egocentric: if an agent's representation of a scene is incomplete relative to their own, infants use the agent's representation *to predict* and *interpret* the agent's actions (see for a review Baillargeon et al. 2016, p. 171). Infants keep track of what events an agent has or has not witnessed in a scene, and they attribute appropriate epistemic states to the agent: they expect an agent who has witnessed an event to *know* about it. As Peter Carruthers (2013, p. 147) pointed out, there is a crucial *background inferential premise* whereby «seeing *leads* to knowing» [my italics]. If Sally sees that P, then she *knows* that P. Therefore, mindreading «system contains a computational rule which generates the conclusion, [Sally knows that P], as output when provided with, [Sally sees that P], as input. Here the knowledge that seeing leads to knowing remains implicit in the inference rule in question» (Carruthers 2013, p. 147, Carruthers 2017). This premise bases the inferential chain that links contents information under a *propositional format* (see below § 1.4 and Chapter 5 § 1.2) from a very early developmental stage.

Following the representational TOM perspective, I will argue that the ability to metarepresent other beliefs is the fundamental feature of the mindreading, whose capacities are many and very sophisticated, but the main one is supposed to be just *the capacity to represent via propositional format others' perspective about a shared referent to predict their behaviour*. Predicting here means *attributing* knowledge to someone. Such attribution entails an expectation that generates a prediction. Let's see in detail what a representational TOM implies.

1.1 Being metarepresentational creatures

Representational TOM makes infants understand mental states as representations (e.g., Caron 2009; Gopnik&Wellman 1994; Leslie 1994), and then that such representations are equal to the representations about the world (Kovács et al. 2010). In other terms, there would be a unique mechanism that allows infants to represent the state of the world and the others' mental states. Therefore, young children recognise that

an observed agent holds a belief about the world that cannot match their own belief or the state of the world. In this way, they take into account others' point of view about something whose perception is shared. On the basis of this common experience, like for example the location of a toy into a box or behind an occluder, infants are able to infer others' future behaviour accordingly with the change of the environmental constraints maintaining distinct from the others' expectation their own representations about a given scenario. Therefore, infants compute what they see - all the experiments I am going to illustrate involve visual perception (but see Träuble et al. 2010[2]) -, and at the same time they compute what others see or cannot see about the same referent.

The crucial point I would like to emphasise is that the representations about others' perspective *influence* the very infants' representation about the world (Kampis et al. 2013; Kampis et al. 2015; Kovács et al. 2010). By the notion of *belief ascription* here, I intend a type of computation about others' perspective that occurs spontaneously and it regards informative contents inferred by the observation of the scene. The concept of *belief* here is to interpret in a broad and weak sense, i.e., as informative content that can be epistemically vague and opaque but it must be semantically definable (Recanati 1997) in a propositional format, e.g.: "the ball is inside the box", or more simply "there is something there". These simple propositional representations are formed swiftly and efficiently.

Following Kovács' (2016) model, I will attempt to show that the infants perform the spontaneous belief attribution (termed by the Hungarian researcher *Belief-file*) much more flexible than it was supposed to be so far. Nevertheless, during the development, the early «representational capacities do not alter in any fundamental way», as suggested also by Carruthers (2013, p. 142) in his defence of Leslie et al.'s (2004) mindreading account. The hypothesis is that «what matures over time are the quality and extent of the interactions between this system and executive, attentional, and planning mechanisms» (Carruthers 2013, p. 142; Carruthers 2015; Jacob 2016; Kovács 2016; Marchetti&Sangiuliano Intra 2015; Tirassa et al. 2006), that enriches its ray of action. The flexibility of belief attribution processing is favoured by the existence of a

[2] The participants (15-month-olds) at the experiment (whose aim was to test early belief understanding in different belief-inducing situations), had both visual and «manual information access as a proper basis for belief induction» (Träuble et al. 2010, p. 434).

unique mechanism enabling to represent the world and the others' beliefs: the computation of a false belief warrants infant observers to infer something about the surrounding world. This allows a rapid shared knowledge of the world.

In this way, the primary form of mindreading *may support the bias of universality assumption predicted by natural pedagogy theory*. However, possessing the innate representational capacity does not involve the existence of a full mindreading system, rather the necessary premises of just a «single mindreading system that exists throughout, but which undergoes gradual conceptual enrichment through infancy and childhood» (Carruthers 2015, p. 141).

1.2 The nativist account does not imply full-fledged mindreading at birth

Recent striking and revolutionary findings (Buttelmann et al. 2017; Krupenyean et al. 2016[3]) suggest that great apes possess to some extent a rudimental understanding that an agent's action is grounded on her beliefs about reality. The hypothesis is based on an implicit false belief test which 12-month-old human infants pass; the researchers adapted it for chimpanzees, bonobos, and orangutans. The great apes, as a group, distinguished true and false beliefs in their helping behaviour. Furthermore, great apes are able to engage pretend plays like human babies (Gómez 2008; Pellegrini&Smith 2004). Although we cannot claim that non-human primates possess mindreading system enable to engage social interactions among peers and with environment like human beings (Tomasello et al. 2003), this discover may represent a further sustain to the existence of an innate – even if not mature - form of mindreading. Moreover, it is compatible with the prediction that humans can develop mindreading in virtue of an interaction «with other cognitive systems, such as working memory and executive control» (Jacob 2016, p. 1).

Therefore, mindreading may be seen as an output of a complex cooperation among several human cognitive mechanisms underlying pragmatic competences, and

[3] Using the anticipatory looking (AL) false belief test (as Southgate et al. (2007) methodologically developed), Krupenye and colleagues found that chimpanzees, bonobos and orangutans are able to anticipate the location where an actor (disguised as their peer) would search for an object based on her own (false) belief about object's location.

executive functions, working memory system, and eventually the competitive cooperation with natural pedagogy too. In this view, on the one hand, the full-blown mindreading onset is the result of an uneven phylogenetic route that may leave the subject individual times and modes (relatively influenced by social and cultural groups contexts cues) to emerge and perform. On the other hand, the ontogeny of specific folk-psychology capacities is innate and it can be described experimentally and sufficiently divided into generalizable developmental stages that should be approximately the same throughout different cultures (around the preschool years) (e.g., Callaghan et al. 2005; Liu et al. 2008; Sabbagh et al. 2006)[4]. Therefore, it is impossible to underestimate the variations in «developmental patterns for individual children and cultural groups that testify clearly to the influences of particular inputs and interactive social experiences on the progression of theory of mind development» (Wellman&Peterson 2013, p. 66). Such variations are nicely illustrated by the differences between children in Western cultures (e.g., USA, Australia, and Germany) compared to Chinese and Iranian children. But, at the same time the universal onset of mindreading capacities is undeniable in a normal developmental trajectory. This is compatible with a modularistic account of TOM as described by Scholl and Leslie (1999) (see below Chapter 5, § 7). Furthermore, the possibility to establish a developmental time table has important implications for clinics. Indeed, analysing the lack of specific mentalization abilities at given development stage represents a crucial marker that indicates the presence of neuropsychological disorders, like autism and William syndrome[5], or late-signing deaf children (Peterson et al. 2005; Peterson 2009; Wellman&Peterson 2013)[6].

[4] Callaghan et al.'s study (2005, p. 378) used a standard procedure «to measure false-belief understanding in five cultures: Canada, India, Peru, Samoa, and Thailand». They found «synchrony in the onset of mentalistic reasoning, with children crossing the false-belief milestone at approximately 5 years of age in every culture studied».

[5] For what regards new insights about the relation between impairment representational TOM and autism and William syndrome see Kampis et al. 2017; Senju et al. 2010.

[6] It has been attested theory of mind delay among deaf children of deaf parents (Wellman&Peterson 2013).

1.3 Against nativism

I propose *to reject the radicalism of the constructive account*, according to which the capacity to attribute false beliefs is taken to be the result of «a cultural process tied to language acquisition» (Perner&Ruffman 2005, p. 214). Young children would be able to pass implicit tests following an associationist strategy (according to an hypothesis sketched for the first time by Perner and Ruffman (2005) and updated by Heyes (2014)), whereby infants are supposed «to form three-way associations between an agent, a toy and a location, and to expect these associations to persist in space and time» (Hemling et al. 2016, p. 441). In the behavioural-rule account, infants perceive agents acting on objects, but their expectations are regularised by statistical patterns. In other words, infants should learn very soon behavioural rules about how agents normally perform in specific situations. Thereby, infants apply such rules to interpret and predict agents' actions (e.g., Perner 2010, Ruffman 2014; for a review see Baillargeon et 2016, p. 177). De Bruin and Newen (2012) even postulate an *association module*, that enables infants «to register congruent associations between agents and objects», and an *operating system* «which allows them to transform these associations into incongruent associations through a process of inhibition, selection and representation» (De Bruin&Newen 2012, p. 240). The ontogeny of mindreading is predicted to be later and strictly connected to social context and later sophisticated cognitive skills like the use of language (de Villier 2005; Garofoli&Fenici 2017). In this view, language is not only necessary for children to make progress in their understanding of false beliefs, rather - as Lohmann and colleagues claimed (2005, p. 245) - «children come to understand beliefs as a result of participating in various kinds of linguistic interactions». However, there is a disagreement about which aspect of language is so crucial for mindreading development. Lohman, Tomasello and Meyer (2005/2012) indicate that conversational practices about deceptive objects and training on the syntax of complementation elicit 3 year-olds' false-belief understanding. Other authors like Montgomery (2005/2012) emphasise mental state vocabulary. In this regard, de Villiers (2003) argues that the grammatical form ("he/she thinks that …") is necessarily required for thinking about belief. Language elicits also major sensitivity to take into account a variety of perspectives through different modalities (perceptual and

discursive ones). Heyes and Frith (2014, p. 1358) proposed that explicit mindreading is «a culturally inherited skill», like literacy, and «it is passed from one generation to the next by verbal instruction». In this view, reading minds is permitted by the same cognitive process underlying reading words ability, whereas the implicit and automatic competences to ascribe false belief are due to neurocognitive mechanisms that allow yielding proper expectations about other's behaviour.

On the contrary, here I accept Grice's work heritage based on the idea that verbal comprehension is a form of mindreading. Grice considered «the comprehension process as starting from a metarepresentation of an attributed utterance and ending with a metarepresentation of an attributed thought» (Wilson 2000, p. 412). As Sperber did sum up: «Verbal understanding consists *in forming a metarepresentation* of a representation of the speaker» [my italics] (Sperber 2000, p. 133). When speakers intend an utterance and hearers interpret it, they metarepresent it, i.e., the mentally represent the utterance as bearer of specific content. Following this line of thought, natural languages incorporate a full-fledged meta-representational capacity, whereby we can claim that natural languages *presuppose* meta- representational capacity, and not the inverse (Sperber 2000). Linguistic interactions do not require only the recognition of intentional agents (as suggested by Lohman et al. 2005), but also more sophisticated representational skills.

1.4 Nativism vs. constructivism on the propositional nature of attributed beliefs

The very full-blown mindreading rests on metarepresentational resources, that are present from the early infancy. This is the nativist core idea I embrace. *Nativist approach do not deny that culture and experience make the difference and improve the performances*. Therefore, the TOM scale designed by Wellman and Liu (2004) for pre-schoolers - that maps the development progression through several «tasks tapping different aspects of understanding persons' mental states»[7] (Wellman&Liu 2004, p.

[7] The researchers tested 75 children aged 3 - 5 on all of tasks including diverse-desires task (DD), diverse-beliefs task (DB), knowledge/perceptual-access task (KA), false-belief task (FB), and a hidden-emotions task (HE).

523) – cannot support constructivist account since all stages of TOM scale require «a grasp of the representational nature of the mind» (Westra&Carruthers 2017, p. 166). For this reason, I will linger over the implications of the representational formats that I suppose to be propositional also in preverbal infants. *Propositional* means *conceptual*. In this regard, I will tackle briefly Davidsonian's triangulation account, and endorse Michael's (2015) suggestion to adopt the notion of *unicept* following Millikan (2013) (Chapters 5, § 1.4, § 1.5).

Propositional entails also the possession of *aspectuality* skills. This issue is the core of the argumentation *against nativist account* either for researchers like Apperly, Butterfill, and Low - who use to distinguish between an early-developing system (System 1, or *minimal mindreading*) that allows tracking rudimental forms of mental states, and a later-developing system (System 2) that is based on a full development of belief concepts and other propositional attitudes[8] -; or for Perner and colleagues (as Rakoczy, Huemer, Leahy, Fizkte and others) who suggest that it is impossible for children younger than 4 years of age handling spontaneously the notion of aspectuality[9]. Perner and Roessler (2012, p. 519) have defended the view that early capacity to hold false beliefs shown in implicit tests reflects «unconscious social knowledge of lawful regularities». According to them, the standard false belief task «requires explicit consideration of the agent's subjective perspective on his reasons for action». This implies a capability to capture «an intentional switch of perspectives» that it is supposed not to be possible before 4 years of age. In their view, to pass direct tests, infants need to interpret agent's behaviour as an intentional action, namely an act performed for a reason, and then, they need to understand the difference between objective reasons (i.e., usually non-psychological considerations that justify doing something) and subjective reasons (i.e., subject's perspectives on her own objective reasons), which requires «awareness of different perspectives» (Perner&Roessler 2012, p. 520-521). Furthermore, in more recent studies, Perner and colleagues (2015; Huemer et al. 2017;

[8] For more details see Chapter 5, § 2 where the two-systems account is discussed and criticised.

[9] Following a known example provided by Recanati, by the term "aspectuality" we have to intend that skill which enables to represent simultaneously without contradiction that Superman and Clark Kent are the same person (Recanati 2012, pp. 196-198). I will tackle the notion of *aspectuality* and its implications in Chapter 5 § 4.2, § 4.3, § 4.4.

Perner&Leahy 2017) have mobilised and accommodated Recanati's theory of mental file for providing an explanation of such impossibility, which, however, in my opinion, is grounded on ideological assumptions that attempt to force the interpretation of experimental results (see Chapter 5, § 4, § 4.1). On the contrary, *belief attribution machinery seems to show very early a great flexibility in several tasks also those encompassing dual identity object recognition*, as we will see (Chapter 5, § 4.4, § 4.5) thanks to recent findings made by Buttelmann et al. (2015) and by Kampis (2017). According to their results, infants distinguish the double identity of the objects, and such accomplishment enables them to use person's belief about the object identity in order to infer his/her goal and to help his/her accordingly.

To sum up, according to *reach* nativist interpretations (e.g., Baillargeon et al. 2010; Kovács et al. 2010; Leslie 2005; Southgate et al. 2007; Kampis 2017), the core *conceptual* capacity needful for representing belief as such operates already in early childhood. Against this *rich* account, *lean* accounts (e.g. Perner&Ruffman 2005; Zawidzki; 2011; 2013) deny that children represent beliefs before 4 years of age, and provide deflationary explanations. A halfway theoretical construct has been offered in these last ten years by Apperly and Butterfill (2009; see also Apperly 2011; Butterfill and Apperly 2013; Low et al. 2016) who put forward the abovementioned two-system account of mindreading, that I will discuss in more details in the next Chapter (§ 2). Nevertheless, even if we accept a precocious metarepresentational system based on a propositional attitude format, we could not explain the following developmental gap: why do infants pass implicit false belief task before the first birthday, and fail verbal explicit false belief test until the age of 3-4? At the end of Chapter 5, I will provide two kinds of answers. One concerns the striking recent neuroscience findings comparing the performances of infants, children, and adults relative to the recruitment of social brain networks during simple false belief tasks. The results emphasise no significant developmental changes from infancy to childhood and to adulthood. The second answer, instead, follows Helming, Strickland and Jacob (2014; 2016), who hypothesise that the same basic pragmatic-communicative features that are predicted by natural pedagogy theory lead to misunderstanding also the verbal request in a standard false belief test. In this regard, I will also mention the similar Westra and Carruther's (2016)

proposal concerning the power of referential bias that – according to them – would not interfere with mindreading process triggered in third-person view (Chapter 5, § 6.3).

To conclude this long introduction that lists all the themes I will unfold in the present chapter and in the subsequent Chapter 5 dedicated to early representational format of mindreading, my intention is to make evident the cooperation between natural pedagogy theory and infant primary form of mindreading, at least from two aspects. Firstly, the extreme flexibility of belief attribution processing makes available the universality assumption for the benefit of natural pedagogy; then, the communicative dyadic interactions bias to some extent the comprehension of others' instructions in non-pedagogical contexts. The progressive maturation of the mindreading system inhibits the pedagogical instances. Mindreading becomes necessary for social learning in more sophisticated context where children have to «differentiate trustworthy, benevolent, and reliable communicative sources of information from communicators who are unreliable, uninformed, or downright bad intentioned providers of useless or deceiving information» (Gergely et al. 2007, pp. 145-146). The relation between the increasing capacities to evaluate reliability of informative source and the decrease of natural pedagogy stance will be argued in Chapter 6.

2. False belief computation as TOM signature

It has been researched for forty years, and the most common way of investigating TOM has been through the use of false-belief tasks (Premack Woodruff 1978; Wellman 1990; Wimmer&Perner 1983). False belief task has been considered as a litmus test for sketching TOM in terms of representational capacities. We have seen that in a typical explicit false belief task, the child observes an agent who places an object in a container A and then a displacement phase begins with the agent leaving the scene and the object is transferred by another agent to a second container, B. The test is aimed to establish whether the children realise that the agent erroneously believes that the target-object is still in location A. This have been tested by asking the children *where* the agent will look for the object (Baron-Cohen et al. 1985; Wimmer&Perner, 1983). The results obtained by this experiment repeated in a variety of ways have led to deem until 2005

that children younger than 4 years cannot hold false beliefs (for a review see e.g., Baillargeon et al. 2016; Birsch et al. 2017; Scott et al. 2010; Wang&Leslie 2016; Wellman et al. 2001). Only at that age, thus, children would «begin to realize mental states such as beliefs are not direct reflections of reality, which must always be accurate, but representations, which may or may not be accurate» (Onishi&Baillargeon 2005, p. 255; see also Callaghan et al. 2005; Wellman et al. 2001; Wimmer&Perner, 1983).

2.1 The introduction of implicit false belief tasks and the new experimental paradigms

Starting from 2005, Kristine Onishi and Renée Baillargeon suggested that a representational theory of mind is present much earlier than 3-4 years of age, and that preverbal infants are equipped with such ability. To test this hypothesis their investigations used new experimental paradigms: the spontaneous-, or implicit-response tasks. This choice was due to make easier for the infants the comprehension of the standard verbal false belief test relying not on their verbal and executive competences, but only on their putative capability to infer false beliefs. The term *implicit* refers to «reasoning that occurs without conscious awareness of the processes involved but can be accompanied by awareness of the contents generated» (Baillargeon et al. 2016, p. 178). In other words, the experimental challenge is that whether infants hold false belief, then such capacity should be manifested through their spontaneous reactions produced while they observe in third person perspective particular scenes involving an agent and a switch of object's location. Spontaneous-response tasks include violation-of-expectation (VOE), anticipatory-looking (AL), and active helping paradigm (He et al. 2011, p. 292). VOE tasks have the aim to verify whether infants look reliably longer at an agent acting inconsistently with her false beliefs. The experimenters measure the duration of gaze attention focused by the infants to the inconsistent choice as opposed to the consistent choice performed by an agent following her false beliefs. The VOE method was adopted by Onishi and Baillargeon's «groundbreaking» research (2005) that represent historically the first spontaneous-response task with infants presented to the scientific community. Instead, AL tasks investigate whether children manage to anticipate where an agent will search for the object following the false belief about its

location. In this case, the experimenters have to determine the direction of infants' gaze. One of the first experiment that adopted AL with a non-verbal false-belief test was conducted by Southgate, Senju and Csibra (2007) using an eye tracker to measure anticipatory looking. Finally, Buttelmann, Carpenter and Tomasello (2009) arranged the active helping paradigm to test false belief understanding in infants. They examined whether 18 month-old infants helped an adult achieve his goal, assuming that in order to comprehend that goal «infants had to take into account what the adult believed» (Buttelmann et al. 2009, p. 337)[10].

2.2 The problem of 'replicability'

It is worth noting that some very recent studies doubt the reliability of prior pivotal tests about the measure and the investigation of mindreading capacities in early infancy (Burnside et al. 2017; Crivello&Poulin-Dubois 2017; Kulke et al. 2017; Kulke&Rakoczy 2017; Phillips et al. 2015; Powell et al. 2017; Priewasser et al. 2017). The attempts of replication regard some of the best-known experiments, structured accordingly to the three abovementioned paradigms: VOE, AL, and helping task. For example, Kulke et al. (2017) tried to replicate some AL-paradigm based findings, as Southgate and collaborators' (2007) study. Kulke and colleagues failed a full replication because their results were only partially congruent. Another failure regards the replication of Kovács et al.'s (2010) conducted by Phillips and collaborators (2015), who did not manage to provide further evidence for automatic TOM in adults. For my argumentation Kovács and collaborators' (2010) finding is very important, as I will unfold below (§ 3.1). Part of this meaningful finding is grounded on the comparison between infants and adults' automatic false belief computation relative to an agent completely irrelevant to the task performed. Kovács herself with other colleagues

[10] Active helping paradigm has been adopted also in many other experiments, see for example Buttelmann et al. (2014) based on object location, or Buttelmann et al. (2015) study which shows that 18 month-olds can help successfully an adult «to achieve her goal based on the adult's belief about an object's identity». Also Scott and Baillargeon (2009), using VOE paradigm, tested the early false belief attribution capacity to discriminate object identity. These studies suggested how infant's false belief attribution might be sophisticated and flexible. It is to be noticed the fact that different methodologies arrive at the same conclusions.

(2014) provided confirm for spontaneous belief attribution in adults in analogue contexts of Kovács et al. (2010), using neuroimaging technique (see for more details Chapter 5, § 5). However, Phillips and colleagues (2015) argued that some procedural modifications adopted in this last finding (i.e., Kovács et al. 2014) may support their hypothesis. Therefore, as it is evident from this brief report, the debate about replication procedure has just begun and deserves further clarifications. For instance, the case of Powell et al.'s (2017) conceptual replications. Powell and collaborators tried to replicate indirectly some experiments as Onishi and Baillargeon (2005) (based on VOE paradigm as we will see in the next paragraph) with procedural differences and changes, and «failed to reproduce evidence that 18-month-old infants' expectations are violated when an experimenter who ought to have a false belief reaches for an object in its true location» (Powell et al. 2017, p. 4). However, Powell and collaborators «found evidence consistent with the original reports» for what regards Knudsen and Liszkowski (2012), and Buttelmann and colleagues' (2009) study based on active helping paradigm (see also above footnote 10) (Powell et al. 2017, p. 9).

Taken together all these attempts show that a robust 'replicability' is only given partially and relatively to some test phases and not to the whole experiments. Although some specific findings imply serious difficulties to be replicated, and this fact makes decrease the scientific robustness of data interpretation, here I assume the general reliability of the remarkable multitude of studies that - conducted so far through different methods - involved and compared infants, children and adults with the scope to test their representational and inferential competencies in social interactions.

2.3 Onishi and Baillargeon's «groundbreaking paper»[11]

Onishi and Baillargeon (2005) investigated whether 15-month-olds were surprised to see an agent act incoherently with her true and false belief about an object's location. During three familiarization trials, infants could see a toy (a watermelon slice) between two boxes, green and yellow one. Then, an agent opened the doors in the back wall of the apparatus, and after playing with a toy for a while, hide it in the green as opposed to the yellow box, and she leaved. When she came back, she reached the toy from the

[11] In these terms Carruthers judged Onishi and Baillargeon study.

green box and paused with her hand inside the box. The trial ended with a curtain that was lowered in front of the apparatus. The infants and the agent know that the toy is surely inside the green box, and at this point infants received four belief-induction trials. In the knowledge-green condition, the agent observed the yellow box moved alone a very short distance and then returned to its original position. The toy was still in the green box, and infants ascribed to the agent such true belief. In another true-belief condition (yellow-condition), the agent watched the toy moving from green to yellow box. This time infants think the agent truly knows the new location of the watermelon toy.

«In the test trial, the agent reached into either the green or the yellow box and then paused. In each condition, infants expected the agent to act on the information available to her, whether it was true or false» (Baillargeon et al. 2016, p. 174). According to the researcher's predictions, infants should not be surprised when the agent chose the green box in the first condition, and the yellow box in the second one. In addition, the results confirmed this expectation comparing the different looking time spent by the infants for the inconsistent choice performed by the agent when she searched for the toy inside the yellow box in the former condition, and inside the green box in the latter. Indeed, only during the agent's inconsistent act, infants increased their gaze attention because their expectations have been violated. When the researchers compared these data with the two following false belief conditions, they predicted to obtain the same results. In the false-belief (FB) green condition, the agent was absent when the toy moved from the green box into the yellow box. Therefore, if infants were able to keep separate what they know through their own visual perspective from what they believe the agent can know about the change location occurred during her absence, they would expect the agent to reach the toy into the green box. In other words, infant should ascribe a false belief to the actor about the novel location of the watermelon in virtue of the fact that they know that the agent did not see the toy movement. The computation of such belief is operated in virtue of the hypothesis that infant take into account the visual perspective of the agent. In the same manner, infants should infer that the agent hold a false belief in the more sophisticated FB yellow condition, in which the agent saw the toy moved into the yellow box, but once she leaved the apparatus the toy turned back to the green box, i.e. its original position. In this case, in fact, infants looked

longer when the agent reached the toy into the green box, i.e., when their expectation was violated. Briefly, on the basis of this experiment, it is reasonable to claim that:

> «infants (i) distinguished between their own and the actor's perceptions; (ii) kept track of what the actor did and did not see; and (iii) understood that the actor's perceptions (rather than their own) should be used to predict her behaviour» (Onishi&Baillargeon 2005, p. 257).

Fifteen month-olds appeared to predict other's behaviour not simply on the basis of her ignorance about the state of things, rather on the basis of a computation of a false belief. Indeed, they responded appropriately even when the agent was mistaken, and not simply ignorant, about the watermelon's location. In 2007, Surian, Caldi and Sperber adopted VOE paradigm for investigating 13 month-olds' capacity to attribute false belief and have expectations even about a nonhuman agent's future actions on the basis of the agent's «exposure to relevant information about the object's location» (Surian et al. 2007, p. 581). Their results converged with Onishi and Baillargeon (2005) study suggesting more precisely that infants had no specific predictions about the agent's action when significant information was not perceptually available to the agent (a caterpillar searching for a cheese slice), and had not been accessible before the action (Surian et at. 2007, p. 584).

Other investigations based on AL method attempted to provide further support to dispel any doubts concerning ignorance or false belief attribution emerged in such VOE experiments. For instance, in 2007, Southgate, Senju and Csibra remodelled the Onishi and Baillargeon's experiment measuring the anticipatory gaze of 25 month-olds while they watched a video in which a puppet bear hid a ball in one of two boxes, and then a female actor (whose only the head was visible upon a panel) reached through one of two windows behind the boxes to retrieve the ball. In the test trial, after having seen the ball being hidden in the left box, the actor was distracted by a phone ring and turned away from the scene (Southgate et al. 2007). When the phone stopped ringing, the agent turned her face toward the boxes, and the windows lit up. Most infants managed to correctly anticipate the agent's behaviour, and looked at the window above the left box, i.e., where the agent falsely believed the puppet to be hidden (Southgate et al. 2007; for a review see also Scott et al. 2010, pp. 367-369).

These experiments contradict the alternative association-interpretation put forward by Perner and Ruffman (2005) in order to explain the results achieved in AL and VOE tests. An associative cognitive process is too weak to account for the complexity of mental attitudes ascription that involves at least two kinds of mental states: motivational and epistemic/informative ones. Leaving aside the intriguing and fundamental question connected to the infant capacity of attributing motivational mental states (such as desire, goals, dispositions) (e.g., Leslie et al. 2005; Steglich-Petersen&Michael 2015), the focus remains on the other kind of mental states, also termed «the reality-congruent informational states» (for a review see Scott et al. 2010, p. 368), that corresponds to states of knowledge (and lack of knowledge). Knowledge attribution to agents comes about through perception (i.e., what the infant sees from the visual perspective of the agent), through memory about what the agent has seen, and through inferences (i.e., what infant thinks the agent reasonable infers from the scene). The critical issue that has been investigated so far is whether infants are capable of ascribing to an agent a representation of a scenario that is distinct from their own perspective. As Baillargeon and coworkers (2016, p. 171) have summarised, ten years of developmental psychology research in this field indicate that at least from the second half of the first year of age, «*infants may use the agent's representations to predict and interpret [her] actions*» [my italics]. Before the first birthday, infants have the striking ability to keep track of what an agent (it does not matter if human or not) has witnessed in a scene, and on the basis of this competence they ascribe properly epistemic states to the agent. Such attribution is manifested (and it can be measured) through certain degrees of expectation: infants expect an agent who has attended an event to act consistently with the knowledge gained. Therefore, the capability to take into account the agent's knowledge from agent's visual/spatial perspective (VSP) represents the initial and crucial point, the fundamental hinge that allows epistemic states attribution.

3. Taking into account other's perspective

I assume here that taking account VSP is an innate spontaneous cognitive process that represents a building block for the primary form of mentalization (Freundlieb et al. 2017; Kovács et al. 2010; Samson et al. 2010; Senju et al. 2011; Surtees et al. 2013).

Furthermore, from evolutionary perspective and consistently with the most recent findings about TOM capacities in nonhuman primates (see above), the understanding of someone else's point of view is shared by other non-human animals (Freundlieb 2017; Hare et al. 2000; 2006 for the case of chimpanzees[12]). In Samson and colleagues' study (2010), for example, the participants at the experiment automatically processed the content of an avatar's perspective, even if it was not relevant for their task. Specifically, the researchers showed that participants spontaneously computed the number of objects that the avatar could see. Their results indicated that the adult participants were faster on those trials in which their own perspective was consistent with the avatar's perspective rather when it was not, suggesting that adults are able to use automatically efficient processes to compute what others can see (Samson et al., 2010; Freundlieb et al. 2017). Freundlieb, Sebanz, and Kovács (2016), through five experiments, showed that participants adopted spontaneously VSP of another agent, as long as she was interpreted as an intentional agent.

3.1 VSP-taking and mentalization in infants: «susceptibility to others' beliefs»

A posteriori, such intriguing outputs may constitute the theoretical basis for another important and turning point finding on infant's false belief ascription ability, conducted by Kovács, Téglás and Endres (2010) with the participation and the comparison both of adults and 7 month-olds. In their paper, *The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults*, Kovács and colleagues (2010) reported more precisely seven experiments which compared the behavioural response of adults and very young children in an object-detection task where an agent was present or absent in the scene but he was never involved directly in the task. Specifically, the adult participants saw different animated movies where a humanoid blue puppet put a ball on a table and he reached the left part of the screen from where he could observe the subsequent scenes in a similar visual perspective of the experimental subjects. The ball could move by itself and go behind an occluder located on the table.

[12] Chimpanzees, ravens (Bugnayar 2011) and other mammalians (Kaminsky et al. 2005) show great capacities to take successfully VSP of others into account in competing interactions with conspecific for food.

Sometimes the ball stayed behind the occluder, and sometimes it escaped far from the screen. The experimenters created four conditions with the intent of manipulating the formation of two true belief conditions, and two false belief conditions in respect of the presence or the absence of the ball behind the occluder. In the first true belief condition (a), the ball rolled behind the occluder and stayed, whereas in the second true belief condition (b) the ball went away. Both the agent and the participants saw the ball's movements, so they created the "same" belief about the ball's final location. In the first false belief condition (c), the agent left the scene and in his absence the ball moved out the scene, whereas in the second false belief condition (d), participants and the agent saw the ball that left the scene, but when the agent left the scene too, the ball returned behind the occluder. At this point, the agent's belief diverged from the participants, but his belief was irrelevant for the task. Indeed, the «participants were not required to perform belief computations» (Kovács et al. 2010, p. 1832), but just to press a button as soon as they detected the ball behind the occluder in the final trial, when participants saw that the agent returned and the occluder was lowered showing the presence of the ball. The results indicated that the adults were faster when both they and the agent hold the first true belief (i.e., all observers believed that the ball was behind the occluder). Participants were slower in the second true belief condition, considered as the baseline condition in which neither the adult subject nor the agent believed the ball to be behind the occluder (*Ibid.*, p. 1832).

Regardless of the task request, even if the participants computed the agent's false (or true) belief, such representation should remain independent of their own perspective and their knowledge about ball location. In this manner, the adult reaction time response should not manifest any correlation to the (true or false) belief of the (intentional but passive and) irrelevant agent. If representing others' beliefs implied only to store others' informative contents (which are not referenced to the current state of the world), then such representations should not affect how persons interact with the surrounding environment. The amazing results of this study contradict such prediction, suggesting more sophisticated functions for false belief computations. Indeed, when the researchers compared the baseline condition with the c) false condition (in which only the agent believed that the ball was behind the occluder), they found that the adult participants were faster than in the baseline condition. The results ought to be the same

instead, because participants both in baseline condition and in c) condition did not expect to see the ball. On the contrary, the experimental subjects were faster both in c) and in d) than in the baseline condition, and moreover, there was no significant difference of the reaction times between c) condition and d) false condition in which only the participants knew that the ball was behind the occluder. According to the authors, this suggests at least two things: the first one is that «*just seeing* the agent automatically made participants compute his beliefs and that these beliefs were represented and sustained similarly to participants' own beliefs»; the second one is that participants «computed the agent's belief and that this belief *influenced their behaviour* even though it was inconsistent with their own belief» [my italics] (p. 1832).

Kovács and colleagues (2010, p. 1832) wondered also if the computation of false belief exert its influence on the participants only in the presence of the agent to whom the false belief is ascribed, or it is maintained «in parallel with their own beliefs». For attempting to solve this question, the researchers tested a new group of participants who watched the same video of the prior experiment with the only difference that in the final trial the agent was replaced by a pile of boxes that compared into the scene just before the occluder was lowered. The results of this experiment were similar to the first experiment: «participants were faster than in the baseline condition when they and the agent (who was not present when the occluder was lowered [because he was replaced]) believed the ball to be behind the occluder» (Kovács et al. 2010, p. 1832). In false belief condition, like in the prior experiment when only the agent believed that the ball was behind the occluder, reaction times were faster than in the baseline condition. This means that participants' behaviour was influenced by agent's beliefs although he was not present when they detected the ball. Therefore, we can claim that «adults automatically compute and store the beliefs of other agents»[13]. In order to investigate whether very young infants (fifty-six 7 month-olds) may have analogue capacities to represent false beliefs, and to be influenced by them, Kovács and colleagues (2010)

[13] A third experiment was designed to demonstrate that participants computed beliefs and they were not distracted or influenced by perceptual factors and habituations. For this reason, participants saw the same videos of prior experiments but without the agent. A pile of boxes was present in every movie in the place of the blue animated puppet. In this way the participants did not recognise any intentional agent and did not attribute any beliefs. Participants demonstrated to be faster detaching the ball when they knew the ball was behind the occluder regardless of the pile of boxes.

designed four experiments using VOE paradigm and they followed the model of the prior experiments. Indeed, the movies involving the blue puppet were similar to the previous experiments but conceived to generate the expectation of finding the ball behind the occluder, although no ball was found when the occluder lowered. The reaction times were measured in terms of infants' surprise by the duration of their looking. In the first of the four experiments, infants showed to be more surprised, and thus they looked longer when both they and the agent «believed that the ball was behind the occluder» (but they did not find it), rather than when «neither the infant nor the agent believed that the ball was behind the occluder» (baseline condition) (Kovács et al. 2010, p. 1833). Crucially, the researchers compared (in a second experiment) the baseline condition with the false belief condition «where only the agent believed the ball to be behind the occluder». When the occluder lowered and no ball appeared, even if infants expected to find no ball, they looked longer in this condition rather than in the baseline condition where both the agent and the young participant knew that no ball was present. Kovács and collaborators (2010, p. 1833) hypothesised the possibility that infants expected the agent's surprise, anyway, the results indicate at least that infants computed «the agent's belief and looked longer when this belief was not confirmed». Like adults, also infant's reaction was influenced by the computation of the supposed agent's beliefs, even though such beliefs «clashed with the infants' own beliefs» (Kovács et al. 2010, p. 1833). In a further experiment[14], the researchers - as they did with the adults' sample - tested «whether infants would maintain others' beliefs even in the agent's absence». For this purpose, they presented infants the same movies of the second experiment just mentioned with the comparison between the baseline condition (in which both the agent and the infants expected to find no ball behind the occluder) and the false belief condition (where only the agent expected to find it). The only difference was that at the end of the trial the agent did not come back, but a pile of box entered the scene. The results confirmed the infants' surprise like in the prior false

[14] Kovács and collaborators (2010) designed another experiment to guarantee that no visual differences among the movies could affect or overwhelm belief reasoning. For this reason, they presented a sequence of movies in which at the end the occluder did not lower and they compared the infants' reaction between the baseline condition, and the false belief condition. The results showed no different looking time behaviour. This suggests that no surprise was triggered, thus, in the other experiments they did actually compute false and true beliefs and not mere visual perceptive differences.

belief experimental conditions, i.e. they looked longer at the scene in false belief condition (even if the agent was not present at the end) rather in the baseline condition. This suggests that infants, like adults, can make online appropriate computations of others' beliefs and they are able to sustain them even when the agent is absent. Consequently, at least two crucial aspects of (implicit) false belief reasoning can be posited on the basis of this finding:

i) the *representations about others' beliefs* seem to have similar properties to the *representations about the environment*, and the former, like the latter, can affect adult and infant behaviour.

ii) The simple presence of a social agent may be enough «to *automatically* trigger online belief computations [my italics]» both in adults and just in 7-month-old infants (Kovács et al. 2010, p. 1834).

In conclusion, to sum up, what the experimenters found is that «subjects were automatically computing the beliefs, and false beliefs, of the observer, and that these third-person expectations were priming (or in this case inhibiting) the subjects' own action systems just as their own expectations did» (Carruthers 2013, p. 165). Even though the infants expected the absence of the object when the screen lowered and it reveals actually that there is no object, «the presence of a violated expectation attributed to another agent makes this event seem interesting (and makes the infants look longer)» (Carruthers 2013, p. 166).

In the next two paragraphs I am going to illustrate two experimental candidates that can provide respectively a behavioural confirm (Senju et al. 2011) and a neuroscientific one (Parise et al. 2015) for i) and ii).

3.2 *VSP-taking and mentalization in infants: visual access and consequent beliefs attribution in 18-month-olds*

In Senju et al.'s (2011, p. 878) study, 18-month-old infants received a first-person experience with a visually opaque or trick blindfold that looked like opaque but it was transparent. Subsequently, they watched a movie in which an actor wore the same

blindfold (opaque or trick one), and meanwhile, an object was removed in front of her. By using AL paradigm, the researchers «assessed whether infants own experience with the blindfold influenced their prediction of where the actor would search for the displaced object» (Senju et al. 2011, p. 878). Assigned randomly to Opaque condition (OC), or Trick condition (TC), infants first familiarized with the visual properties of the assigned blindfold for five minutes. Toys were presented, and an experimenter asked directly: "Where is the [object]?". In the test phase infants observed a video sequence in which a female actor «sat behind a panel with two windows, in front of which were placed two boxes» (Senju et al. 2011, p. 879). When a small toy appeared on one of the boxes, the windows lit up and simultaneously a ring sounded. These were the signals for the actor who reached after a 1750-ms delay through the window toward the object. In other two trials, a puppet entered the scene and put a toy into one box, then closed it and left. When the windows lit up, the ring sounded, and after the same delay (1750-ms), the actor searched for the corresponding location of the toy. Instead, in the last trial, first the puppet collocated the toy inside the left-hand box, and then the actor wore the identical blindfold previously worn and experienced by the infant. At that moment, the puppet took the toy away leaving the scene with it, and then the actor removed the blindfold, the windows lit up, and the ring sounded.

In the two conditions, infants showed different expectations about the actor's action after object shift. While in OC they expected the actor to look for the toy in the left box, in TC infants did not reveal such an expectation[15]. These results strongly suggest that in OC infants attributed a false belief to the female agent, and by contrast, in TC they did not ascribe any belief about a specific object location to the agent, because they knew on the basis of their own previous visual experience with the trick blindfold that she had been able to see when the puppet had removed the toy from the left box. These results induce to claim that «infants used self-experience with the

[15] The authors tracked the direction of infants' first gaze following the windows' light, and the differential looking score, that was obtained by subtracting total looking time to the right window from total looking time to the left window and dividing the outcome by the total looking time to both windows. In OC, most of the eighteen infants made first gaze towards the left window, while in TC just six infants looked at that direction. This suggests that the target of their first gaze was conditioned by infants' previous experience with the blindfold. The looking score remarkably differed between the two groups (Senju et al. 2011, p. 879).

blindfold to assess the actor's visual access, updated her belief state accordingly, and used this attribution for action prediction» (Senju et al. 2011, p. 879-880).

These results represent a further proof that infants did not understand such «false-belief scenarios by a reliance on learned behavioral rules» (Southgate 2013, p. 10), as proposed by Perner and Ruffman (2005) because they never saw the agent wearing the blindfold, so they could not acquire any behavioural correlation. Therefore, infants solved this false belief scenario by representing the world from the other's perspective.

3.3 Neural signatures in preverbal infants for representations attributed to others

Kampis, Parise, Csibra and Kovács (2015) measured gamma band activity over the temporal lobes (by EEG technique) in 8 month-old infants involved with sustained object representations after occlusion event. The aim of the authors was to compare the gamma band activity when infants observed an object occluded, or when the object simply disappeared, simultaneously with another agent observing similar events from another perspective. Through the oscillations of gamma-waves, the researchers wanted to track the putative infants' computation of agent's representation observing an identical scene where an object was occluded (compared with an object disintegration event) from different visual perspectives. If infants are able to take into account an agent's visual perspective and subsequently to compute her mental representation about shared events, then such computation has to be *visible* through specific brain activity. For this reason, Kampis and coworkers designed their experiment using a previous finding accomplished by Kaufmann, Csibra and Johnson (2003), who reported the neural correlate of object permanence in 6 month-olds. Specifically, they noticed «a burst of gamma-band EEG activity over the temporal lobe that occurs during an occlusion event and when an object is expected to appear from behind an occluder» (Kaufmann et al 2003, p. S140). They interpreted such increase as corresponding to the infants' mental representation of the occluded object. Therefore, Kampis and colleagues used Kaufamann et. al.'s methodology and they assume that the infant's computation of an occluded object (namely, the infant's sustained object representation) is manifested by the burst of gamma band activity. On this basis, to test the infantile capacity to attribute mental representations (e.g., sustained object representation) to others, the

authors predicted an analogue EEG response in those cases where infants observe an agent seeing an object occlusion (or an object disintegration). In other terms, they investigated neural responses when infants observe an agent who has to sustain the representation of the same occluded object. For this reason, the researchers designed a familiarisation trial in which infants observed on a screen the target object placed on a rotating panel-box with two sides removed. On the left of the panel, a female agent looked in front of her. The panel box turned around itself revealing the object. Then the agent turned her gaze toward the target object, and in this way, for several seconds, both the infant and the agent saw the same thing but from different perspectives. The box started to spin again, and initially the toy became occluded only for the agent, and not for the infant. After another portion of time, the object was gradually occluded also from infant's perspective due to the rotation of the panel-box. Kampis and colleagues (2015) predicted that the occlusion from the agent would have triggered a sustained object representation also for the infants. Indeed, they measured the average EEG gamma-band activation (around 25–35 Hz) over the left and right posterior temporal regions during the occlusion both of the object from the actor's first, and then the single infant's view. This striking result confirms the prediction. In fact, when there is the crucial passage from the segment of time in which both the agent in the movie and the infant see the target object to the condition in which only the infant can see the object due to the partial box rotation, no particular gamma band activation should occur if infant did not compute any mental state attribution. Only once the panel box finishes its rotation, the infant – theoretically - should represent the permanence of the object beyond the box, and subsequently the gamma-band activity should increase. However, this experiment contradicts such scenario. Indeed, even though infants were facing the object when it was hidden from agent's perspective, they made register a burst of gamma band activity as such as when, a few seconds later, the object was going to be hidden from their own perspective. This entails that 8 month-old infants «successfully computed the visual perspective of the actor regarding the object», and they attributed to the agent «the representation of the continued existence of the object behind the occluder» (Kampis et al. 2015, p. 5). However, the experimental conditions described so far regard the true belief held by both the agent and the infants about the persistent existence of the target object after the occlusion. As further and crucial proof of an

early metacognitive competence, Kampis and colleagues (2015) tested whether so young infants were able to ascribe to the agent also a false belief about the object's permanence behind the occluder. For this purpose, the researchers developed an experiment with the same familiarisation phase, except that, in the test trial, when the target object was faced in front of the infant and simultaneously hidden to the actor, the object suddenly disappeared. «This disintegration was therefore visible to the infants but not to the actor; hence, this event must have resulted in the actor's false belief that the object was still behind the occluder» (Kampis et al. 2015, p. 5). The alleged attribution process of such false belief should be implemented during the disintegration event, and thus, it should be indicated by an increase of gamma-band activity. This prediction was confirmed by the results which compared the false belief condition with the true belief condition in which the disintegration was visible both to the agent and the infant. In false belief condition, the authors noticed the same average gamma band activation (25-35 Hz) of the prior study.

What does such study show us? In this study, infants demonstrate to infer that another subject keeps on representing the object, despite the sharp conflict with their own visual perspective. This is strongly evident above all in the false belief condition, where «the infants must have encoded that the other person had seen the object being occluded, but did not see the disintegration, and hence the attributed object representation could not be discarded on her behalf, but had to be possibly refreshed and sustained further» (Kampis et al. 2015, p. 6). Therefore, this finding may constitute a further solid confirm that before the onset of language, infants can already hold a metarepresentational comprehension of other minds. «By possessing such powerful representational capacities, infants are endowed with the ability to ascribe to others any representations they themselves can form, including representations that are in conflict with their own representation of reality» (Kampis et al. 2015, p. 7).

In respect of Kovács and colleagues' (2010) abovementioned study, this experiment provides confirms from a neuroscientific perspective about the attribution belief process automatically triggered in very young infants. Consistently with Kovács et al. (2010), Kampis and colleagues put forward that «the cognitive systems involved in representing the world from infants' own perspective are also recruited for encoding others' belief» (Kampis et al. 2015, p. 1), as predicted by Kovács et al. (2010).

For my purpose, two are the most important insights that emerge from Kovács and colleagues' experiment (2010), and they regard: 1) the relevant recruitment of false belief reasoning for learning something about the world through others' perspective; and 2) a conceptual interpretation of others' behaviour based on propositional attitudes. Given the close connection between *world-representations* and *others' false beliefs-representation*, and the influence employed by the latter on the former, infants may gain information about the world from (false) belief attribution practise in specific circumstances. Precocious infants' use of false belief attribution reveals a lot about the very nature of belief "handled" by infant's mind. The most evident function of false belief attribution is to *predict others' forthcoming actions* thanks to the integration of different mental states, as for example goals and beliefs, and the relative consequences of these states for action interpretation[16]. Therefore, given particular constraints, false belief computation serves to achieve information about the world, and by this acquisition to predict others' behaviour. The latter issue has been explored by Kampis, Somogyi, Itakura and Király (2013) through VOE methodology applied in a false belief preference task involving 10 month-olds; while the former has been investigated in very young infants (6 month-old) by Southgate and Vernetti (2014) using EEG technique, with the intent to verify whether so young infants can appreciate the causal relation between beliefs and action.

3.4 Neural indication for very young infant appreciation of belief causal role in action interpretation

Under certain conditions, the looking-time paradigm cannot indicate precisely whether infants expect that «a particular belief will lead to a particular action outcome» (Southgate&Vernetti 2014, p. 2). Nevertheless Kovács and colleagues (2010) have shown the infants' sensitivity to others' belief, it is unclear whether infants of that age (around 7 months) «have any appreciation of the functional consequences of those beliefs» (Southgate&Vernetti 2014, p. 6). Furthermore, when infants are very young, it

[16] As it is known, such connection is central to Dennett's "Intentional Stance", that allows to explain and predict the others' behaviour (Dennett 1987).

is difficult relying on anticipatory eye movements because infants are not so able to visually disengage from a stimuli when they are too young (6-7 months of age), as Elsabbagh and colleagues (2013) have shown. Instead, action prediction can be measured through the analysis of neural motor cortex activation using EEG technique, as it is well known and attested both in adults and in infants (Marshall&Meltzoff 2011; 2014; Southgate et al. 2009; Southgate&Begus 2013) by tracking the amplitude variations of alpha band activity (or mu rhythm desynchronization) during the observation of others' action[17]. Measuring the different degrees of alpha mu suppression value in 6 month-olds, Southgate and Vernetti (2014) found that when an agent has a false belief about the location of an object, motor cortex activation indicates that infants are able to predict if the agent will reach for. Specifically, if the agent falsely believes that, for example, 'the ball is in the box', infants predict that she is going to reach for the box, but if the agent has the false belief that 'the ball is not in the box', they do not predict she will act at all. In the false belief condition where the agent believes the ball to be present but the infants, who are observing her on a screen, know the ball to be absent, the presence of motor activation suggests that young participants expect the agent to reach for the box, and trigger their action prediction on what she, rather than themselves, believes to be the case. This amazing result reveals a very important fact: that infants can «appreciate the causal role that beliefs play in action» (Southgate&Vernetti 2014, p. 2). This manifests their capacity to link specific beliefs to related actions, and, thus, to «understand the functional implications of beliefs for action» (Southgate&Vernetti 2014, p. 7; Dennett 1987).

Another point emerged from this research is worth highlighting: such finding does not demonstrate any causal role of motor/mirror system for mental state attributions, as already indicated by Jacob and Jeannerod's (2005) argumentation. Rather, Southgate and Vernetti state further that the mirror system is «unlikely to be involved in the process of attributing mental states» (Southgate&Vernetti 2014, p. 9). In fact, both false belief experimental conditions presented and compared - namely, the first condition in which the agent falsely believed that the ball was in the box, but infants knew it was absent; and the second condition in which the agent believed the ball was not in the box while the observers knew it was – should recruit obviously mental state attributions.

[17] For more details see chapters 1 and 2.

However, according to the results, only in the first condition sensorimotor cortex activation has been recorded. This fact suggests that «the involvement of the motor system is one of predicting *how* the expected action will unfold» (Southgate&Vernetti 2014, p. 9).

4 «Learning about the world through the lens of others»[18]

Kampis and colleagues (2013) started from a theoretical framework whereby the infantile false belief reasoning is a very useful social practise for learning some aspects about the world that infants can reuse in order to interpret other's behaviour, and thus *to predict* their future actions. The researchers found that the false belief attribution is not limited to that agent involved in the contingent scenario that infants observe. Furthermore, belief attribution does not regard only the mere preferences or dispositions that single agent expresses toward a specific object. It is emerged that being involved in a true-false belief scenario allows infants (at least from 10 months of age) to achieve much more information that is not restricted to the understanding of the specific person's mental representations. Kampis and colleagues aimed to overturn the person-specific interpretation that characterises above all those false belief tasks based on a preference choice between two objects, from e.g., Woodard (1998) to Lou (2011) experiments. Following the suggestions put forward by Gergely et al. (2007) and Egyed et al. (2013)[19], Kampis and colleagues (2013, p. 233) sustain an *object-centred* approach whereby in referential situation false beliefs attribution may induce infants to learn «new information about the referent (about the particular object that was referenced)». Furthermore, false belief reasoning would allow infants to extend what they understand to other similar situations, going beyond the contingent condition the information achieved belongs to. In this view, the epistemic content is treatable and extendible to other people and not only attached to the person who is referring to it. Under another guise, it seems that the main characteristics of the universality principle (predicted by natural pedagogy theory) can be found very early in referential but non-pedagogical contexts, as the experimental conditions of Kampis et al. (2013) study show. I am going

[18] Kampis et al. 2013, p. 237.

[19] See chapter 3 for more details about these experiments.

to present the Kampis et al. (2013) study in order to motivate better its relevant implications about this issue.

The researchers (Kampis et al. 2013, p. 234) designed an experimental setting very similar to Luo's study (2011) setup with the intent to investigate whether infants, who observes an actor expressing her attitude toward a target object, would «expect a newly introduced agent to have the same attitude». By doing so, they wanted «to test whether 10 month-old infants would encode knowledge conveyed in a social situation as person-specific, or if these situations trigger the acquisition of more general knowledge (in a non-person specific way)» (Kampis et al. 2013, p. 234). Infants were presented a series of videos in which they could watch a female agent A who demonstrated an attitude preference towards an object (e.g., on her left) rather than another one located on the right in front of her. Each one of the two objects was behind a transparent screen. There were two conditions, whose the crucial one was, of course, the false belief condition, where an occluder (a read cardboard) avoided A to see the object placed on her right, that was removed by a hand invisible to A. In this way, A falsely still believed that there were two objects while the infant saw that there was only one[20]. The hypothesis was that infant inferred that A believed to make a real choice when she grasped the object remained on her left. Instead, in the other condition, there was no occluder intervention, and both the agent B and the infant could see the removal of one of the two objects. So, B demonstrated to grasp the object, but this act was not interpreted as a choice.

The authors argued that the false belief condition represented «the strongest test of the person-specific encoding» (Kampis et al. 2013, p. 234). In fact, in the second phase, the infants of the first group, who had seen A being involved in her false choice, could now watch B in front of the same objects seen before by A but located reversely. If the person-specific view were correct, infants should not expect any preference since they are observing a new agent. However, the results contradict such prediction. Indeed,

[20] The red cardboard (collocated next to the transparent screen) created an opaque screen that avoided the agent to see the object. «During removal, a hand reached behind the screen and removed the object from the scene. This step was invisible to Agent A in the occlusion conditions but visible in the no-occlusion (control) condition. During the test phase the location of the objects was switched, and another agent (Agent B) was introduced» (Kampis et al. 2013, p. 235).

through VOE, the researchers measured an increase of looking time duration when B made an inconsistent choice in respect of A's preference. In other words, infants expected that B demonstrated the same preference attitude shown by A.

Kampis and colleagues (2013) compared this degree of expectation with VOE measures in the no-occlusion condition where infants could now watch the novel agent (A) involved in a choice towards one of the two objects. The new agent A made her choice, but no particular surprise was noticed both in the consistent and inconsistent choice performed by A in respect of the previous B's preference. In this case - it has been argued -, infants could not have any expectation, because they had not been attributed any belief to the agent B, who actually did not make a choice[21]. Briefly, when in the occlusion condition infants computed the (false) belief about A's preference, they attribute such selective preference not only to A but also to other subjects as well. The authors suggest that the belief content attributed by infants to others does not refer to an agent's motivational factor, rather to an epistemic content regarding the referent object. In other words, through the computation of false belief, infants infer some information about the object that they reuse in the subsequent situation applying to other persons.

For my purpose, the great importance of this study consists in the suggestion that the automatic taking into account the other agent's knowledge serves «to predict not only the actions of this agent exclusively, but also the actions of other agents» (Kampis et al. 2013, p. 238). Kampis and coworkers (2013, p. 238) argued on the basis of their results that «infants did not handle the emergent information based on someone's visual access in a person-specific way. They have used the acquired information to predict the

[21] As a further control condition, in another group of infants, the researchers presented the same identical scenario but during the test phase in the occlusion (false belief) condition they introduced as a novel agent a 2 year-old child instead of the female agent B. The results confirmed a longer looking times during inconsistent events. However, in order to exclude all the possibilities that infants expected a choice from agent B identical to agent A because they could not distinguish properly B from A, Kampis and colleagues conducted a further control study. Following the procedure designed by Buresh and Woodard (2007) to explore 13 month-olds' capacity of distinguishing several agents involved in goal-directed action, Kampis and coworkers (2013, p. 237) created two conditions where they did not shift the object location. In the first condition, A remained in the test phase but wearing a different t-shirt. In the other condition, B appeared during test phase, and chose like A the same object (with the objects location still being unchanged). The results revealed through the looking time patterns «a novelty preference for Agent B, suggesting that infants were able to distinguish between the two agents» (p. 237).

actions of other agents accordingly». The infant's inference regarding the salient agent's visual perspective - as it is configured in the false belief condition - does not refer to the agent's motivation (or the preference for itself, e.g.: "A prefers the yellow cylinder") rather to the target object. The referent is the hub of the computation, namely *the referent representation is the epistemic content of the attributed (true or false) belief*. Belief attribution proceeds accordingly to the computation of others' visual perspective, and then infants use «the inferred mental states in their evaluation and prediction of forthcoming actions and their outcomes» (Kampis et al. 2013, p. 238). In their experiment, for example, the referent object has been interpreted and categorised through the false belief attribution as a *good object*; then, such information has been supposed to be true/valid for everyone, namely it has been ascribed to conspecifics. The false belief reasoning drove infants' interpretative efforts toward the referent target of the agent's choice rather than toward the subject's motivational disposition, accordingly to the object-centred interpretation proposed for pedagogical conditions (Gergely et al. 2007; Egyed at al. 2013). However, here there are no-pedagogical or communicative ostensive conditions that would trigger any universality assumption. The infants observed the scene in a third person view, and the samples examined were younger than those ones in Gergely and coworkers' experiments. The Kovács et al. (2007) study together with Kampis et al.'s (2013) results indicate how the infant representational mind works in presence of one or more agents involved in a social context within a scenario experienced from different points of observations. Furthermore, these experiments show how infants can benefit from different social conditions for learning through others about the surrounding environment. For this purpose, early false belief reasoning serves several aims, but it works with object-centred stance. This does not mean that toddlers are not able to adopt a person-specific interpretation during preference task, but probably before the first birthday the object-centred cognitive stance is predominant until, for a certain period and under particular circumstances (like pedagogical contexts, see Egyed et al. 2013), «both strategies are available in parallel» (Kampis et al. 2013, p. 239). Kampis and co-operators (2013, p. 239) suggest that by the maturation of full-blown mindreading abilities person-specific belief encoding becomes predominant.

The flexibility of early mental representation attribution reveals its powerful social adaptive advantage in the lack of binding between belief-holder and the conveyed epistemic content, which can be reused by infants as «a shared knowledge […] applicable to other agents as well» (Kampis et al. 2013, p. 238). Furthermore, how the authors suggest, «the lack of binding of mental states to agents […] could serve an important role in promoting joint action and cooperation» (Kampis et al. 2013, p. 238). Therefore, such precocious form of ascribing epistemic content may constitute the cognitive basis for the omniscience bias occurred in pedagogical relation where another kind of commitment is required.

5. Conclusions

Under the light of the epistemic principle, the necessity to learn quickly and efficiently in a variety of social contexts (into which infants are constantly immersed) can make their mind so flexible to benefit as many circumstances as possible. The salience information obtained from a false belief condition can represent in this way a precious information to use in further social interactions. I deem that such general tendency to attribute universally epistemic contents can be accommodated for more sophisticated social learning context, as pedagogical contexts actually are (where, in fact, infants must first understand communicative cues, then the actions directed toward the functions of an object or toward the application of social norms for subsequently imitate them and demonstrate to have learnt). In this way, the natural pedagogy system makes use of this innate capacity already in place between 7 and 10 months of age, and employed in other kind of social interactions.

CHAPTER 5.

Mindreading and assumption of universality. Part 2) Propositional nature of attributed beliefs

1. Problems about the nature of attributed beliefs

The early TOM features that I sketched in the previous chapter do not imply necessarily the existence of a double mindreading system, as long as we assume a nativist point of view about the notion of belief whereby the infant's capacity to understand and attribute (true/false) beliefs prescribes the propositional nature of the very beliefs. If we assume this, then it becomes acceptable the existence of only one mindreading system that changes during the first years of a normal ontogenetic development through passages that have to be interpreted «as points on a continuum rather than discrete categories» (Kampis et al. 2013, p. 238). Given the complexity of such assumption and the different representational levels infants must integrate[1], more conceptual clarity about the nature of such kind of beliefs is necessary, even because the hot debate on theory of mind faculty turns around the very nature of beliefs that infants can sustain from a certain developmental stage. In fact, if infants are not immediately able to engage well-formed propositional attitudes, a double mindreading system may constitute the best explanation for infants' success in implicit false belief tests. If propositional attitude represented a later cognitive acquisition, what it seems to be a precocious onset of a TOM's primary form should be explained in other ways.

Which is the kind of attitude infants are able to ascribe to others? Kampis and colleagues (2013, p. 238) are inclined to guarantee the propositional nature of early infant belief attribution against the alternative proposals provided by Rakoczy (2012), and by Apperly and Butterfill (2009). The former has problematised the issue about the regular form of beliefs, which must respect at least three characteristics: *inferential integration*, *accessibility to consciousness*, *conceptualisation*, and he prefers to adopt the notion of «subdoxastic states» for early epistemic states attribution. Apperly and

[1] The infant's object representation is primarily influenced by the VSP taking account of the agent; then, such affected object representation (i.e. , the informative content about the referent) is ascribed not only to the agent but also to other people.

Butterfill (2009) have proposed a two-system model for mindreading based on a primitive form of belief called *belief-like*, supposed to be as more suitable and primitive for the initial stage of cognitive development. Kampis and colleagues (2013) claim that the beliefs attributed by infants should not be distinguished from proper beliefs, even though they recognise that the emergency of a full-blown TOM system is a later achievement. Here I defend what Kampis and colleagues advance and what Kovács (2016) has attempted to systemise about the cognitive structure for the early belief attribution. I will start to argue that, on the bases of the experimental findings reported so far, it is reasonable to accept that the three fundamental characteristics indicated by Rakoczy (2012) for proper beliefs are satisfied by infant representational cognitive structure, even though in a weak sense.

1.1 The weak sense of belief concept in infancy

Following Rakoczy's (2012, p. 61) description of regular beliefs[2], beliefs' inferential integration means that beliefs are «inferentially promiscuous», or in other words, beliefs combine other beliefs that produce inferentially further beliefs. Beliefs combine also «with other attitudes in practical reasoning to yield decisions and actions» (Rakoczy 2012, p. 61). Thanks to Southgate and Vernetti's experiment (2014, p. 2) we have seen how very young infants may appreciate «the relationship between beliefs and other mental states, or beliefs and action». Therefore, it seems that the first feature of regular belief is quite satisfied. Rather, it is to be noticed that even if action prediction can be executed accurately by 6-7 month-old infants, it can be the outcome of the computation only of the others' perceptual experiences, without infant necessarily being aware that what the others are holding is a belief (Southgate 2013). Putting in a nutshell, if infants are capable to represent an event from the others' perspective, it is not so clear whether they understand that what others are representing from their perspective is a particular mental state that can be defined as *belief* (Rakoczy 2012). This is the issue concerning with the features of accessibility to consciousness, and the notion of conceptualisation as "believing about believing". The following passages may still belong to a sub-personal, non-conscious level: *i*) possessing a representation about an

[2] Rakoczy in turn mentions Stich (1978) and Davies (1989).

event, *ii*) taking into account the representation of another agent about the identical event, and then *iii*) manifesting a certain sensitivity to this representation.

However, according to Baars (1988; Baars et al. 2003), who proposed the global broadcasting cognitive architecture for consciousness, and also according to the more recent Carruthers' ISA theory (2011)[3], «conscious states are those that are widely available to other systems, especially those involved in belief-formation and decision making» (Carruthers 2009, p. 122). In this regard, the active *reuse* of such distinct representations for action prediction, or for helping intervention, for example, constitutes, in my opinion, a further step towards a form of awareness. Therefore, I reject what Rakoczy claimed following Block (1995, p. 231), that: «from a practical point of view, the clearest evidence of access consciousness of a state is the person's ability to use the state in the rational control of speech (i.e., to express and *self-ascribe the beliefs in question*) and the best evidence of inferential promiscuity is the ability to *reason verbally* in promiscuous ways» (Rakoczy 2012, p. 64 [my italics]). The clearest evidence does not imply that it is the only one. Preverbal infants manage to handle actively contrasting representations for action prediction and for learning something about the world, as it is emerged clearly in the particular false belief conditions designed both by Southgate and Vernetti (2014) and by Kampis et al. (2013), and this does not involve any kind of speech, or assertive claim that preverbal children could never declare. The fact that preverbal infants lack the personal-level capacity to have beliefs about beliefs in terms of self-ascription - namely to have a (conscious) higher-order belief with the content: "*I believe* that Sally will go to the box" - does not entail

[3] The Interpretive Sensory-Access (ISA) theory proposed by Carruthers (2011) is a very complex version of the self-other parity account concerning with the nature of self-knowledge. In his view, human beings have an inner direct access only to the mental representations whose origin is merely sensorial; all the rest of knowledge about our own propositional attitudes is a matter of interpretation. Specifically, the British philosopher states that knowledge of our own thoughts is always the result of a self-interpretation process based on the same sensory channels that we use when working out the others' mental states. In this view, the metarepresentation of the first-person is hierarchically subordinated to the mindreading faculty that at a certain stage point starts to swift its activity towards own Self. This implies other problems because, as Marraffa and Meini (in press) have highlighted, if metarepresenting the Self is a metacognitive activity, not every metacognitive process is metarepresentational. Furthermore, even though ISA theory outlines a single phylogenetic and metarepresentational route for mindreading and metacognition, it does not manage to predict that mindreading should also be developmentally antecedent to introspection.

that infants do not have the expectation that Sally will go to the box. The experiments mentioned so far attest the formation of such expectation based on foregoing beliefs which are combined and integrated inferentially until the formation of an expectation. A package of information gained through perceptual access and linked each other inferentially, as in the following scenario in which an infant sees that:

- "The toy is in the box" –
- "Sally sees too that toy in the box" –
- "The toy is in now in a new location" –
- "Sally have not seen the change location" –
- "Sally does not know the new location" –
- "Sally wants to reach the toy", lead to the final expectation:
- "*Sally will go to the box to reach for the toy*".

Such expectation may be considered in turn a belief definable through the propositional content: "Sally will go to the box", that is the outcome of an inferential chain of simple epistemic contents endowed with propositional form that intertwines according to the principle of rationality. Such cognitive process implies an inferential integration that drives infants toward the final outcome which is manifested not only through a surprise (calculated by means of the gaze duration), but also through a 'translation' into an active decision making: helping Sally to reach for the toy if requested, for example. In this way, a level of integration and a certain level of consciousness seem to be satisfied for properly accommodating the nature of regular beliefs in preverbal infants, except so far for what concerns the third feature that refers to the bonds between propositional attitude and conceptualisation.

1.2 Representational, propositional and conceptual dimensions for infant beliefs

All the experimental results mentioned so far testimony the striking infants' informational sensitivity, that emerges from how they seek and use information to learn about the world in daily social practice. This innate disposition is founded on a representational mind. As Kim Sterelny (1991, p. 21) wrote:

> «there can be no informational sensitivity without representation. There can be no flexible and adaptive response to the world without representation. To learn about the world and to use what we learn to act in new ways, we must be able to represent the world […]. Furthermore we must make appropriate inferences from those representations».

Therefore, representations must be inferentially linked each other. In addition, I deem, that the *representations of information* - born in those social context depicted in the experimental report - *are propositional like regular beliefs*.

It is normally accepted the definition of belief as a "propositional attitude". As Schwitzgebel (2015) nicely puts it, a *propositional attitude* is the mental state of having some stance «about the potential state of affairs in which that proposition is true [and it is] canonically expressible in the form "*S A* that *P*", where *S* picks out the individual possessing the mental state, *A* picks out the attitude, and *P* is a sentence expressing a proposition». As Millikan (1984) and others philosophers suggest, *beliefs are essentially states that* represent *how things stand in the world* (Schwitzgebel 2015). Therefore, a regular belief is a *propositional attitude*, that means it has, as a proposition, a specific meaning which can be formulated by the typical sentence in our belief-tasks context: "Sally knows that the ball is in the box". As an attitude, regular belief involves a mental stance about the validity of its content (i.e. the proposition), e.g.: "[I see] the ball is in the box", thus "the ball is in the box" is *true*, and the infant, like adults, expects that it *must* be true always for everyone. When Sally sees that the ball is in the box, for infants Sally knows that the ball is in the box. On this premise, infants build their expectations. Following Brandom (2000, p. 158) we can claim that: «propositional contents stand in inferential relations, and they have truth conditions». Propositional contents serve in this way both as a premise and as the result of the inferential process. In force of its propositional form, beliefs involve conceptual content, or in other terms, the explicit form of the propositional stance has implicit in it «the conceptual content of the claim which is articulated by the inferences» (Brandom 2000, p. 19). Following an inferentialist line of thought, «the fundamental form of the conceptual is the propositional» (Brandom 2000, p. 12). In this view, a concept is to be comprehended in terms of *expressing a belief*. Within the notion of *belief*, conceptual, propositional and representational dimensions are linked each other in depth.

In this regard, Brandom (2000) wondered why any mental state endowed with a propositional content should be intended as having representational content. The answer proposed is that the propositional contents have a representational dimension in virtue of their «social articulation». A propositional content of a belief, indeed, may have different significances «from the perspective of the individual believer» (Brandom 2000, pp. 158-159). Therefore, the representational content of a belief reflects its *social* dimension. In addition, since what it is represented propositionally is conceptual, conceptual contents are both *inferentially* and *socially* articulated with a propositional form. From this point of view, the attribution of a false belief has to be intended as a socially triggered propositional content that is conceptually incorrect because it transgresses an inferential rule. If an infant forms the following representations: 1) "The ball is no longer in the red box" – 2) "Sally sees the ball is not in the red box", 3) (through VSP taking account) if she manifests to want the ball, infant knows that Sally will not reach for the ball in the red box. Nevertheless, if Sally reaches for the ball in the red box, she is transgressing an inferential rule (not a behavioural one) emerged during a social interaction. Rationally, the inferential line leads to a contradiction from infant's perspective because her propositional representation is "Sally knows that the ball is not in the red box" (conceptual premise), but she reaches for it in that box. This novel propositional representation is an evident contradiction (propositional conceptual outcome). This transgression deserves a verdict, and the infant just like a judge, through inferential and social parameters, delivers her judgment about what is true and false about others' behaviour. Here, "judgment" has to be interpreted as having a Kantian sense[4]. Brandom emphasised the normative character of concept use recalling the «Kantian normative conceptual pragmatics», and in this regard, he wrote:

> «*Conceptual faculty is the faculty of grasping rules*, of appreciating the distinction between correct and incorrect application they determine. Judgings and doings are acts that have contents that one can take or make true and for which the demand for reasons is in order» [my italics] (Brandom 2000, p. 163).

[4] In his Kritik der reinen Vernunft (1781/1787), Kant intends concepts as predicates of possible judgments. Concepts have the form of rules insofar they fix how something ought to be done according to the rule. Accordingly, representing something conceptually entails that everyone else ought to represent it in the same way.

Such a request for reason is a social practise occurring among humans very precociously. Grasping propositional contents – Brandom (2000, p. 165) suggests - is a sort of «practical mastery of the game of giving and asking for reasons», implicitly (in mere observational context) and more explicitly in interpersonal relations involving social dimension of communicating reasons.

Applying these considerations to infants' belief attribution capacity means to affirm that the representational content of very young infants' beliefs has a propositional form, whereby also a conceptual dimension is to be assigned to the nature of infantile beliefs, but «without predicating *truth* and *falsity* as such [italics added]» that are concepts introduced later without a radical conceptual change (Carruthers 2017, p. 682). Here, operating judgments consists of applying automatically, or at least spontaneously, the transitional rule whereby the young observer passes from the content: "The agent sees the ball going into the box" to the other content: "The agent *believes* (or *thinks*, or *knows*) the ball is in the box". «This will become an attribution of a belief that is false if the ball is moved again while the agent is absent, since this representation will not then be updated» (Carruthers 2017, p. 682). The verbs "believe", "think", "know" are equal in this attributive context because they serve to reflect merely the primary mental representation about the agent's perspective on the actual state of affairs.

What it is not properly synonymous, instead, is the use of terms "*autonomous*" and "*spontaneous*". As Carruthers nicely puts, automatic processes «cannot be inhibited by the subject» (Carruthers 2017, p. 675). On the contrary, spontaneous processes take place «independently of external prompting and explicit (conscious) goals, but they nevertheless depend on implicit goals and hence require executive resources» (Ibidem). Carruthers (2017) suggests that in some circumstances one-mindreading system can genuinely promote automatic computations, whereas spontaneous computations in other contexts. Maybe a VSP taking into account occur automatically, but in any case - I would like to emphasise this point - mindreading system does not encode everything that a perceptually accessible agent can see, rather encoding beliefs occurs for «salient agents about salient events» (Carruthers 2017, p. 681).

The transitional rule exemplified above recalls the rule sketched by Byrne (2005): "P, so I believe that P". He proposed it for his looking-outward self-knowledge model

(see Carruthers 2011, pp. 80-81). If P is the knowledge that one (child) achieves from the recognition of a fact, intended in the cases tackled so far as output, for instance, of a salient preference choice made within a false-belief condition, then, one believes that P. If Sally sees that P, then she *knows* that P, Carruthers (2013) suggested, and he added that «system contains a computational rule which generates the conclusion, [Sally knows that P], as output when provided with, [Sally sees that P], as input. Here the knowledge that seeing leads to knowing remains implicit in the inference rule in question» (Carruthers 2013, p. 147). Knowing that P from the recognition of such different action interpretations constitutes the basis of alleged shared beliefs, i.e. common ground knowledge. P may have an updating content, and it derives not only from facts *seen* in the world, but also from other kind of beliefs ascribed to others, as for the case of Kampis and colleagues (2013) experiments for the preference inducted by the attribution of a false belief. This may be considered as a fruitful circular example of knowledge formation elicited by social interactions. The innate representational mechanism allows putting on the *same hierarchically epistemic degree* both what the infant represents about the world through her own perceptual channels, and what someone else believes about the same referent.

1.3 To be believers infants do not need to be speakers

Grasping propositional content does not need either the possession of truth concept as such, or the involvement of a linguistic dimension. Infants do not wait to become speakers to be believers. Following a simpler version of what it is traditionally called the «language of thought hypothesis» (Fodor 1975; Sterelny 1991; Margolis&Laurence 2005/2011), we can *use* language for exposing and simplifying the representational dynamics with their propositional and conceptual dimensions. However, language (either *mentalese* or folk language) is not responsible for the generation of the representations' sequence that expresses concepts propositionally, bounded each other according to inferential and social *rules*. Language does not establish the rules that, along inferential and social paths, articulate the propositional contents of representations which are generated from the perceptual observation of worldly events involving social interactions. Language does not trigger the kind of

representations and the inferential relation, rather the use of language is intended as a mere instrument that permits us to explain propositionally the informative (conceptual) content of the single representation that is simplified to sentence-like. Such an informative content has the status of a regular belief that can be held and handled by infants for being connected inferentially to others' beliefs and to others' behaviour. Such cognitive representational architecture allows infants to navigate the social world very soon as best as they can, «appraising the environment, explaining new observations, and constructing a shared meaning of the world» (Connors&Halligan 2015, p. 1), through the flexibility of infant belief attribution faculty and its entailments like the universality assumption. If it is true, then, we have to break up the bond between believing and claiming (or asserting), and consider at least the origin of belief systems to be not linguistically dependent as instead sustained by some influential philosophers. For example, Robert Brandom (2000, p. 6) reported Davidson's account, according to which:

> «claiming and believing are two sides of one coin — not in the sense that every belief must be asserted nor that every assertion must express a belief, but in the sense that neither the activity of believing nor that of asserting can be made sense of independently of the other» (Brandom 2000, p. 6).

In fact, in Davidson's view, a believer is the one who interprets the speech of others within a triangular relationship. However, despite this crucial and necessary role played by the language use into Davidson's account, part of his "triangulation" theory about the origin of the beliefs can be fruitful for my purpose insofar as it collocates the very notion of belief within the philosophical framework of psychological externalism, whereby «what determines concepts and contents are […] their relations to the outside world» (Amoretti 2013, p. 50). Such perspective can take into account at least the formation of early beliefs in the daily social relations infants are mostly involved with.

1.4 Davidson's triangulation: the belief notion revisited

By the notion of triangulation, Davidson refers to «the result of a threefold interaction, an interaction which is twofold from the point of view of each of the two

agents: each agent is interacting simultaneously with the world and with the other agent. […] Each creature learns to correlate the reactions of other creatures with changes or objects in the world to which it also reacts» (Davidson 1997a, p. 128). The model of triangulation attempts to clarify the nature of beliefs and how mental contents are acquired through and in virtue of social interactions. Triangulation is necessary to fix the empirical content of beliefs about the external world (anchoring human thought to the world) (Amoretti 2013, p. 50); and to acquire the concept of *objectivity* that is so fundamental to articulate and attribute false beliefs.

The Davidsonian notion of objectivity corresponds, in fact, to «the idea that we may be mistaken, that things may not be as we think they are» (Davidson 1997,p. 129). Accordingly, a propositional thought is objective insofar «it has a content which is true or false independently (with rare exceptions) of the existence of the thought or the thinker» (Davidson 1997, p. 29). To be a triangulating creature one must possess the concept of error (or objectivity), that appears clearly when a particular expectation is not fulfilled. Infants show to have such kind of awareness at least towards others' mistakes. Indeed, in a triadic relation, they can judge if the other person is mistaking, or having a false belief about a referent event. Furthermore, infants demonstrate to share knowledge about the world on the basis of a common perceptual (visual) experience. For this reasons, we would be authorised to consider preverbal infants triangulating creatures of some sort without applying to them the overly restrictive notion of «basic triangulation» evocated by Davidson (1997a), Bar-On and Priselac (2011), or Brink (2004) (for a review Amoretti 2012; 2013)[5] to solve the sharp Davidson's claim whereby, in order to grasp the concept of objectivity, the introduction of language is needed (Davidson 1997b). As Sinclair (2005) noticed, Davidson's triangulation cannot explain the emergence of propositional thought insofar the triangulation requires necessarily to be already conceptually equipped. The very intentional action that ought to be interpreted depends on the concept of belief (and desire)[6]. For this reason, Bar-On and Priselac

[5] Basic triangulation is «a non-cognitive, non-intentional, non-linguistic interactive situation in which two (or more) creatures simultaneously react to common external stimuli and to each other reactions to those stimuli» (Amoretti 2013, p. 52).

[6] As Davidson puts: «Intentional action cannot emerge before belief and desire, for an intentional action is one explained by beliefs and desires that caused it» (Davidson 1997, p. 127).

(2011, p. 130) proposed a middle ground triangulation, and other authors as Brink (2004) and Amoretti (2013) put forward the psychological process of joint attention as the best amendment for triangulation. Joint attention is not a case of basic triadic relation because it requires a certain degree of awareness to be part of a communicative triangle and to understand communicative intention toward a share referent item. However, there is something else to add. A preverbal infant may entertain an almost full-fledged triangulation because, as I am claiming, she infers propositional attitudes about the external referent in virtue of her sensitivity to the presence of the other intentional subject who is determinant to understand the observed event. Thus, can the notion of objectivity be treated more flexibly?

Amoretti (2013, pp. 56-57) emphasises various definitions of the concept of objectivity put forward by Davidson and she individuates «a 'strong' and a 'weak' concept of objectivity. The former involves such concepts as those of belief and proposition, requires the idea of a difference between true and false beliefs, and thus presupposes the concepts of error, truth, and falsity […] The latter can be identified with the minimal understanding that worldly objects or other creatures are external and different from us, as opposed to elements in our own mentality». Thanks to this clarification, I can claim that infants are able to sustain the strong definition of objectivity insofar they assume as a base of their inferential chain processing the fact that if they share with another agent the visual access to the same event (e.g., the location of an object), that event is true for everyone. This assumption generates specific and reliable expectations whereby if they see the agent contrasting such expectations, they infer the agent is mistaking, and her wrong action is guided by a false belief. At the same time, other's true or false belief elicits and influences the infant's inferential process. The social interactive relation can give the contents and influence the deployment of the inferential chains along a coherent array of other concepts which little by little are stored and interconnected. A holistic conception of belief system fixation enters into the process of triangulation, as Amoretti (2013, p. 59) nicely emphasised, and such combination between triangular externalism[7] and holism[8] helps to

[7] By the term "triangular externalism" Davidson (1991) calls his own version of externalism, whose exists a variety of types (De Caro 2011). "Content externalism" (or "semantic externalism") is the thesis that «the contents of an individual's thoughts […] depend on relations that the individual bears to aspects

explain, for example, how two concepts about the same referent can differ each other, or how two individuals can communicate even if they do not share identical concepts.

In conclusion, *preverbal infants are externalist conceptual creatures*. They start being believers of some sort, at least from the second half of the first year of age. Infants show to possess the concept of objectivity because they have expectations about other's behaviour. Expectations that can be fulfilled or disregarded on the base of a normative stance that lies on the assumption of the principle of rationality. Sharing with another person the visual-perceptual access to an event, like the location of an object, infants generate expectations driven inferentially on the base of the intertwined epistemic contents generated by the attended context of observation. Premises and conclusions that perceptual and social environment generate are not linked each other casually but according to *an innate rational principle*. The determination of possible mistakes attributed by infants to an agent in a triangulation setup lies, thus, on an innate rational norm. Davidson himself claims that: «The point is not to identify the norm, but to make sense of there being a norm» (Davidson 2001, p. 7). Constantly in search for information, and equipped of representational mind, infants do not need language to form propositionally simple conceptual contents about items of the world.

Probably, within Davidson's perspective there is no chance to arrive at a full description of the *emergence* of beliefs, as he admitted claiming further to be thankful that he was «not in the field of developmental psychology» (Davidson 1997a, p. 128). The crucial issue is not whether infants and young children are able to represent others' beliefs in virtue of their possess of some particular knowledge about belief like adults who fully master the concept of belief. Rather, we should shift the focus on the very notion of objectivity, because *the notion of objectivity can represent the theoretical background for sustaining false belief attribution*. Objectivity can be interpreted as the indubitable perceptual common ground that establishes the shared certainty of the given event between two subjects at least. Objective is the (visual) perceptual access to an

---

of his physical or social environment» (Bernecker 2013, p. 443). The necessity to be committed in a interpersonal relation projected toward a worldly item to have beliefs makes Davidson's externalism both perceptual and social.

[8] Davidson emphasises the holism of the mental «the extent to which various aspects of the mental depend on each other» (Davidson 1997a, p. 126). At the same time, it is the very holism of the mental that makes the emergence of beliefs so difficult to describe.

event that funds the first hinge of the inferential chain. The infant does not need to understand anything about the act of seeing in order to exercise this ability.

To sum up, infants are able to represent beliefs without conceiving of «knowledge about beliefs that characterizes a mature understanding of the concept of beliefs» (Michael 2015, p. 4). Theoretically, this core idea is in accordance with Michael's teleosemantic perspective borrowed by Ruth Millikan's account about «unicepts» (Millikan 2013).

1.5 Teleosemantic proposal

According to Millikan, "unicepts" are «the basic representational vehicles underlying our abilities to (re-) identify objects, properties, relations or kinds» (Michael 2015, p. 4). As nicely Michael (2015, p. 4) puts it, unicepts are «theoretical tool for capturing infants' and young children's ability to represent and reason about beliefs prior to having full understanding of how beliefs combine with each other and with other mental states in contributing to inferences and guiding action». In Millikan's words:

> «A unicept is a specific individual faculty designed for a very specific purpose, the purpose of collecting and integrating information about some particular thing» (Millikan 2013, pp. 3–4).

Unicept "mechanism" takes, collects and integrates various perceptual stimuli and relates them each other to concern a single thing. In other terms, unicept allows that a variety of input signals collapses into a single item. Thereby, unicept can be conceived as «a source of information about a common distal cause of different proximal stimuli» (Michael 2015, p. 17). In some ways, we can define unicepts as an impoverished version of concepts enabled to provide contents for belief. In this sense, they are seen as building blocks. Millikan's teleosemantic approach amended by Michael (2015) may constitute an alternative to descriptivism relatively to mental contents[9]. Michael's aim is

[9] In contrast to *descriptivism*, alternative proposals have been *causal theories of mental* content that attempt to explain how beliefs (desires, hopes, and in general thoughts) can be about worldly objects,

maybe to accommodate a theoretical insight (born far from mindreading issue) in order to bridge the gap between rich and lean TOM's view. Beside the issue of the representation format, one of the cruces of belief attribution mechanism is the problem of information's integration. The integration of informative cues could not occur without a mechanism enables to store and re-identify mental representation of a referent. In this sense, the two-systems account put forward by Apperly and Butterfill attempts to sketch representational swift between tracks of non-propositional states and establishment of propositional belief structure.

2. Almost like a belief: the minimal mindreading account

Apperly and Butterfill (2009; see also Apperly 2011; Butterfill&Apperly 2013) advanced the proposal about the existence of two systems for mindreading: a precocious system enabling to generate representations characterised by non-propositional states (system1), and a later-developing system capable of representing beliefs as propositionally structured representations (system 2). The former (S1) enables toddlers and very young children to track and reason about what is defined as *belief-like* of other agents, but not beliefs as such (at least until the age of 4 when system 2 starts to be triggered (for much details see also Caron 2009; Michael 2015, p. 2). The latter (S2) springs only when children acquire and "handle" the use of language, and they become capable of advanced executive functions and perspective taking that allow them to mature «a conception of belief as a relation between a subject and a fine-grained proposition» (Carruthers 2013, p. 157). From this point of view, children's failure on explicit false belief tasks «reflects a true conceptual deficit» (Fizke et al. 2017, p. 210). In other terms, the success on explicit false belief tasks will depend on S2, whereas the success on implicit test will require only the involvement of S1, which «has signature limits concerning aspectuality». According to such account, infants lack the notion of *aspectuality*, that should not be required for change-object-location false belief tasks. In respect of change-location implicit tasks, the two-system account predicts that infants

claiming that mental representations are meaningful and linked each other «in virtue of a causal connection between a mental representation and some part of the world that is represented» (Adams&Aizawa 2017, p. 1; see e.g., Dretske 1988, or more recent Rupert 2008).

do not compute any propositional attitude, rather they «operate with relational attitudes only» (Fizke et al. 2017, p. 211).

Let us concentrate on the alleged features of S1 which is grounded on three theoretical constructs, the notion of "field", "encountering" and "registration". The first one refers to a spatial and perceptual proximity but without including psychological states. An agent's field encompasses what her face is directed to, as long as there are no occluders (Butterfill&Apperly 2013, pp. 614–615). This entails that an agent *encounters* an object when it is located in her field. At this point, an agent *registers* the object as being in a specific location and not subsequently anywhere else, and when she successfully performs a goal-directed action on that object placed at that location (Butterfill&Apperly 2013, pp. 619; Michael 2015, p. 9).

Briefly, the early-developing S1 shows the capability to represent another agent as *registering* an object where she last *encountered* it, namely where the object was when it entered agent's perceptual field. According to Apperly and Butterfill, registrations are different from *beliefs* because beliefs are sophisticated representations which have propositional form, while *registrations* indicate a *relation* among three factors: an agent, an object, and a location. A registration is stored in a memory box that maintains the information gained during an encounter through a period in which the agent does not directly encounter the object. Thus, if the object location changes after the first encounter, the registration will not be updated, and consequently it will be false. As someone nicely puts it: «representations of beliefs are metarepresentational, while representations of registrations are not» (Steglich-Petersen&Michael 2015, p. 532). Apperly (2011) deems that, in the beginning, infants have only raw and vague representations of others' false beliefs, and fail to distinguish between the several manners the agent might represent the state of the surrounding world. In this way, Apperly (2011) sustains that there is a *representational discontinuity* between the initial system that appears swift, efficient and which remains unchanged throughout adulthood, and the later developing mindreading system that operates slower, but at the same time results highly flexible[10]. Therefore, what Apperly and Buttefill term *minimal theory-of-*

[10] Carruthers (2015) and Westra (2016b) argued that the processes underlying Level-1 visual perspective-taking tasks seem not to be automatic in all cases, whereas the processes underlying Level-2 visual

mind involves the representation of *belief-like* states, «but it does not involve representing beliefs or other propositional attitudes as such» (Butterfill&Apperly 2013, p. 607). However, although S1 is supposed to work without «employing flexible semantic-executive cognitive processes», Butterfill and Apperly assign to it goal attribution competence (Michael&Christensen 2015, p. 219). In their hypothesis, infants ascribe goals without representing them as related to other psychological states. This may be a strong simplification since researches on infant goal attribution reveal high degree of situational awareness and great sensitivity to interrelations between goals, preferences and beliefs (e.g., Cannon&Woodward 2012; Luo 2011; Luo&Baillargeon 2005; Michael&Christensen 2015).

Furthermore, advocates of the two-systems model of mindreading adopt Fodor's (1983) distinction between modular processes and central processes. The former, according to Fodorian proposal are informationally encapsulated, while the latter underlying belief fixation is isotropic and subject to the confirmation holism. Accordingly, minimal mindreading include informationally encapsulated processes, while full-blown mindreading is subjected to confirmation holism (Apperly 2011). Therefore, following Fodor's view, mindreading in *computationally intractable*, as argued also by Zawidzki (2013), who suggests the phylogenetic priority of what he calls *mindshaping* over mindreading[11].

The challenging question investigated by Jacob (2016, p. 8) concerns «what makes the contents of registrations really different from the contents of genuine beliefs». The flexibility of full-blown mindreading is needed in order to represent the

---

perspective-taking tasks appear to be more spontaneous than described by the two-systems account (see also Jacob 2016).

[11] The notion of "Mindshaping" was originally introduced by Mameli (2001). According to Zawidzki (2013, p. xi), «without mindshaping, none of the other components of distinctively human social cognition—sophisticated language, sophisticated and pervasive cooperation, and even sophisticated mindreading—would be possible». In respect of the latter, for Zawidzki, the attribution of full-blown propositional attitudes would be impossible to have before the evolution of «sophisticated practices of mindshaping aimed at making us easily interpretable to each other» (Ibid., p. xii). For what concerns the intractability of mindreading generated by the holism of belief confirmation, Zawidzki indicates the *aspectuality* of propositional attitudes as the main cause.

aspectuality of beliefs[12], while the representations of registrations (generated during the encountering phase) can be gained by minimal mindreading. The aspectuality of beliefs (and other propositional attitudes) is widely regarded as a reliable sign of the propositional character of their contents» (Jacob 2016, p. 13). The crux, indeed, rests on the notion of registrations which do not establish aspectual epistemic relation between agent, object and relative location. This kind of relationship must be very weak, or «unstructured» (quoting Jacob 2016, p. 24), otherwise it will be considered as propositional.

Denying the propositional nature of infant belief seems to be an ideological attitude grounded on rigid assumptions, rather than a good argumentation founded on theoretical and experimental hinges. On the other hand, admitting the propositional format of the early beliefs entails philosophical dilemmas as I sketched in part. However, nowadays, I think there are several findings that have challenged and overcome Butterfill and Apperly's account on the crucial issue of flexibility and aspectuality of infant's early beliefs. In the next paragraph, I will present Kovács theoretical account of belief-file structured by a precocious flexibility of belief attribution machinery; further experiments sustaining such flexibility suggest also the character of aspectuality possessed by the infantile beliefs about others' mental states. This will constitute for my thesis the argumentative tool for challenging other theorists as Perner and Rakoczy who strongly defend the lack of aspectuality in early infant cognitive representational structure (see also Oktay-Gür et al. 2018 for a review and for the presentation of further experiments with children from 3 to 6 years-old who required an understanding of aspectuality). Firstly, I would like to present briefly the study conducted ten years ago by Song and colleagues (2008) that already challenged vigorously some crucial aspects of two-systems account.

[12] The notion of *aspectuality* of beliefs stands for the dependent relation between belief about an object and «the label under which the object is known to the believer» (Perner et al. 2015, p. 78).

2.1 Song, Onishi, Baillargeon, and Fisher's experiment: updating without encountering

Eighteen month-olds watched a sequence of events involving two agents, a ball, a box and a cup as two containers. The first agent (A1) played with the ball, and then hid it in the box. The other agent (A2) looked on. In A1's absence, A2 moved the ball from the box to the cup, and when A1 returned, A2 said to her: "The ball is in the cup!" (*informative-intervention condition*), or "I like the cup!" (*uninformative-intervention condition*). In the test phase test, A1 reached for both the box (*box event*) and the cup (*cup event*), but in the informative-intervention condition, infants who saw the former event looked longer than those infants who saw the latter event. The reverse result was noticed in the uninformative-intervention condition. These data indicate that infants expected A1's false belief about the location of the ball to be updated when she was told "The ball is in the cup!", but not when A2 told: "I like the cup!". In another test, A2 pointed to the ball's location without saying "the ball is in the box!". Also this time, infants expected A1's false belief to be updated. These results provide further evidence that infants in the second year of life not only attribute false beliefs to agents, but also expect agents' false beliefs to be updated by relevant communications involving sentences or ostensive signals (Song et al. 2008). This study results problematic for Butterfill and Apperly's proposal insofar the infants modulated their expectations about the agent's behaviour by the experimenter's communication, «even though the agent did not encounter the object and did not perform a goal-directed action on the object» (Michael 2015, p. 11).

3. Belief file cognitive structure underlying epistemic attribution

Belief file is a theoretical construal aimed to describe «the core representational skeleton for online ToM reasoning» (Kovács 2016[13], p. 516) enabling to track and update the own representations about others' beliefs. *Belief file* is defined as the «basic representational structure» that operates from the early infancy enabling «implicit ToM processes to store information about other agents' beliefs in a format supporting efficient encoding and updating» (Kovács 2016, p. 510). Its structure is flexibly

[13] The paper was available on line in 2015, but here I always mention the printed version in 2016.

articulated on three horns (the agent as belief-holder, the belief-content, and the referent). Only two are the variables that can be separately updated giving the flexible profile to whole structure: 1) the belief-holder and 2) the belief content. If this weak (or almost free) bond offers many possibilities for updating and attributing processes, at the same time it opens theoretical and unsolved problems at the level of fragmentation of belief contents. Indeed, put the case that Sally believes that object A is in the blue box, *and* object B is located in the red box. Is this a unique belief? Or are two beliefs (Belief 1: A is in the blue box. Belief 2: B is in the red box)? Kovács (2016, p. 518) proposes that even if we consider the prior example as one belief, «it is possible that one piece of information could be changed independently from the other». To attribute successfully belief contents to others, i.e., to mentalize properly, it is necessary to track, to re-identify and recruit belief files which cannot be defined by one single variable (the agent, or the content), rather by the flexible relation among agent and the belief content she initially carries. In fact, the agent may have many beliefs, and the content cannot be always unequivocally determined. Therefore, «the minimal criterion for opening a belief file is the presence of an agent» Kovács (2016, p. 518). The key-function of belief file structure is such representational flexibility of the bond between the variable agent A and the variable content X whereby «A and X can be replaced by various agents and a variety of contents» Kovács (2016, p. 519).

For what concerns belief contents, they can be assumed and attributed through VSP taking into account with the following form: "She/he sees = knows | something is in the box". In this case, the second part of the propositional content about the referent is carried by belief holder: it is represented only thanks to the presence of an agent, but it seems to be not strongly binding to her/him, whereby the propositional content can: i) be updated on line; ii) be simultaneously different or in conflict with infant's own representation of the referent (e.g., the object can change the location only for the infant, or it can reveal to have a double identity); iii) influence infant's own representation about the referent; iv) be attributed to other agents. All the four passages a) occur in a continuous belief tracking resting on online events; b) are triggered by the presence of an agent; c) generate spontaneous behavioural predictions.

3.1. Two crucial values of belief file for learning about the world

Therefore, two are the fundamental values of belief file structure that allow learning states of the world by reasoning on others' knowledge contents. On the one hand, «online belief formation requires a potential agent» (Kovács 2016, p. 526); if the presence of an agent is necessary to permit attribution of propositional attitudes (see Kovács et al. 2010), subsequently belief holder can be substituted by «an "agent–placeholder"» to which attributes the belief content (Kampis et al. 2013, p. 238). This permits to attribute the information achieved and stored to other subjects in further circumstances. Such as flexible attribution mechanism ought to be the base of universality assumption. The innate tendency implemented by the epistemic principle drives infants to *capture* and *generalise* information becoming the shared knowledge that allows navigating more efficiently the social world.

On the other hand, belief file can sustain an *epistemically indefinite content*. It is the case of *empty belief file* where an observer has a belief attribution like the following: "Sally believes that there is *something* (of her interest) in that box". The term "empty" does not indicate «a referential uncertainty», or a «representational vacuum». Belief attribution can establish its full propositional form. However, even in the case of *nothing* instead of *something* we could open a belief file (under true or false condition): "Sally beliefs there is *nothing* in the red box". Put the case that there are two identical opaque boxes, and Sally visually knows which one contains her favourite toy, but the observer does not see what the boxes actually contain. When Sally leaves and Anne (a second agent) enters and switches the boxes, the observer is able to predict Sally's behaviour that will be driven by a false belief, even if the observer does not know what the object exactly is and in which box it is precisely located. When Sally returns and chooses and looks inside one box, the observer will infer that if Sally's expectations are not fulfilled then the object is not there, and Sally has to update her belief. Therefore, it is possible to open a belief file and update the belief-attribution «without even knowing what the content of the belief actually is» (Kovács 2016, p. 520)[14]. It is worth noting

[14] A very similar scenario has been modelled for a false belief test with adults having neuropsychological impairment (Apperly et al. 2004); and with 4-5 year-old children (and great apes) (Call&Tomasello 1999).

that, in this case, the actual state of affairs is inferred by the observer in virtue of an attributed belief, when instead it is usually the opposite: an observer infers others' beliefs on the base of what she sees (and) believes about reality. This example represents one of the best cases where we can learn about the world through what we believe about others' epistemic contents.

> «Belief computation mechanisms not only allow us to infer what other people might believe about the world, but it seems that we can also use these belief inferences to learn new information about the state of affairs» (Kovács 2016, p. 520).

3.2 First-person mental representations cannot vanish

I did talk above about the epistemic vagueness of empty belief, but now I vindicate the semantic clarity of such representations that are regular propositional attitudes (i.e. beliefs) held by the observer. The content of belief is expressed by: "something is in the box". Only apparently object representation lacks, because *something* is given and it is in relation with the proximal object, the box. So, the mental representation about the referent is vague, but not so blind. This conflicts with Kovács' interpretation whereby:

> «The operations involving belief files are not impeded by the absence of *precise first-person information regarding their contents*. In fact, the system permits manipulations with empty belief files, allowing humans to ascribe beliefs to conspecifics based on little or no direct information regarding the content of the mental state» [my italics] (Kovács 2016, p. 509).

In my opinion, the hypothesis that belief files are not necessarily dependent on regular representations of actual objects in the environment does not entail an absence of the first-person event representation. If it were so, the acquired information through belief computation of VSP taking into account would be due to the disappearance of the first person view. While on the contrary, we have true or false belief computation in virtue of the co-occurring presence of contrastive representations that a subject holds

simultaneously. *It is the very simultaneity of opposite perspective that permits the judging operations*. Kampis and colleagues' experiment (2015) about sustained object representation illustrated above indicates that infants compute the absence of the object when it ceases to exist, but when infants are in presence of an agent who is justified to believe the permanence of the object behind the occluder, infants can compute *also* the agent's false belief about its permanence. Belief reasoning enriches and influences first-person representation about the referent object in any case, either when it is not specified initially, or when it dissolves. The inhibition of egocentric view investigated by Baillargeon does not imply the silent of first person operability. The inferential process conducted by the first person can silent the egocentric power of own representations since the computation of others' beliefs springs from the very comparison and from the causal links between the others' and the own representation about events. The relevance of belief computations for social learning rests on the "conflicting" and simultaneous tracking of different perspectives when instead Kovács (2016, p. 523) claims that «belief files can be sustained independently of first-person representations». Kampis (2017, p. 140) nicely puts that «tracking others' beliefs results in holding multiple representations that are highly *overlapping in their content and format* [my italics]». It is the very overlapping that allows the mutual conditioning, in line with Kovács et al. (2010) experiments and further results.

The flexible nature of belief structure tells us how fast and adaptive the updating and attribution processing can be, but we have not to forget that such processes are guided both by environmental changes experienced by the observers, and by the inferential chains that link the representations each other. A belief is inferred as *true* or *false* in force of a *judgment* that is spontaneously computed by the first-person through a comparison with own representations reflecting knowledge about reality. One can reasonably argue that, maybe, attributing and computing content beliefs are separate components (Kovács 2016, p. 525). Decomposing belief attribution processing can be fruitful for understanding the initial and ongoing faculties of TOM throughout the development. Thereby, we may distinguish the starting phase of opening the belief file from the computations of its content, and the bond between belief representations and the corresponding agents; thus, the belief evaluations (true or false), and finally the behavioural predictions upon others' (false) beliefs.

By using the term "belief-file", Kovács is aware of the analogy with the prior Recanati's (2012) definition of "mental file". In this regard, we can acknowledge, as Kampis (2017) did, that *mental file theory* may represent a good interpretative and theoretical tool for understanding the starting phase of opening the belief file insofar mental file belongs to the system of mental representations and refers to them as *modes of presentation* of the object in Fregeanian sense[15] (Pagin 2013, p. 136). Depending on how the subject relates to the object, one generates a mental representation of it. In this way, the referent of such relation is known by the subject *only* from specific characteristics and properties, that can be updated and enriched through subsequent epistemic relations with the target object.

4. Epistemic relations with object: remarks on Recanati's mental files

As Pagin (2013) nicely sums up, the most important properties of mental files are: having a referent, and storing through acquaintance relation information (or misinformation) about the referent. «Acquaintance relations are epistemically rewarding in allowing the subject to gain information from the object. […] A mental file is a kind of repository of information, or misinformation, about the referent» (Pagin 2013, p. 137). Therefore, their primary function is to store information (or misinformation) which corresponds to the properties that the subject *takes* the referent to possess. Information can be deleted, updated while the relative mental file stays the same (Pagin 2013, p. 138). Recanati (2012) writes in this regard that:

> «Files are a matter of information clustering. Clustering takes place when all the information derives from the same source […], and when it takes place, it licenses the integration and inferential exploitation of the information in question. The role of the file is precisely to treat all the information as if it concerned one and the same object, from which it derives» (Recanati 2012, p. 42).

[15] «On the Fregean view, a thought or belief can be *about* a concrete individual (its reference), but what matters to the individuation (or the identity) of the thought's content is not the reference of the singular term, but the sense or mode of presentation of the reference» (Jacob 2014).

Thereby, mental files are containers of information that are useful to re-identify an object at later times so that one «can apply knowledge gained through previous encounters» (Perner&Leahy 2016, p. 496). As already emphasised by Millikan (2000), one of the hardest jobs of cognition is re-identifying individuals, objects, properties, kinds through a variety of channels and conditions, and in this regard, according to Recanti's theory, if I think of a "bear" as a bear and I think of it as an animal (because previously I have learnt it only as an animal through linguistic labelling), I am grasping the same referent by two aspects, namely I have two distinct mental files anchored to the same external entity and linked each other. «When two files are linked, information can flow freely from one file to the other, so informational integration/exploitation becomes possible» (Recanati 2012, p. 43). This is crucial for our discourse about aspectuality because it means that a single object may have different mental files depending on modes of presentation, and the subject can identify an object to be the same through epistemic relations between the files. This kind of linking is termed by Recanati *horizontal*, and as Recanati himself claimed: «from a cognitive point of view, linking is a quite fundamental operation» (Recanati 2012, p. 43). The case of belief ascription to another subject introduces the metarepresentational function of mental files. Mental files serve not only to think about objects in the world, but they have a metarepresentational function: «they serve to represent how other subjects think about objects in the world» (Recanati 2012, pp. 182-183). This entails that in subject's mind an object can be referred through two ways of thinking: «a way of thinking of his own (a *regular file*) and a *vicarious* way of thinking (the *indexed file*)»[my italics], that means that «the subject refers to the object through some other subject's file» (Recanati 2012, pp. 184). In the case of belief ascription, a *vicarious* mental file anchored to «the ascribee's point of view» is opened (Recanati 2012, p. 182). In this regard, Recanati (2012, p. 183) introduces the notion of *indexed file*, i.e., a file that stands for another subject's file about a referent. In a nutshell, the subject S_1 has regular mental files about some entities in the surrounding environment, and in the presence of a S_2, S_1 opens an indexed file that she uses vicariously in order to represent what S_2 thinks about those entities. In Kovács' terms, the S_1's Belief file consists 1) in the perspective relation between S_1 and S_2, that determines 2) the relative contents ascribed by S_1 about what S_2 thinks. In Recanati's terms, S_1 must have previously opened a mental file about S_2, so in

this way the indexed file is to be embedded within S_1's file about S_2. Recanati highlights that «an indexed file cannot be directly anchored to a real object, but only via a regular file to which the indexed file is linked» (Recanati 2012, p. 184). However, for our purpose, the crux is to determine the nature of links between such regular and vicarious files in S_1's mind. According to Recanati, while the transfer of information between regular files flows freely (horizontal linking), in the case of an indexed file linked to a regular one the vertical linking that is established does not allow information to flow freely (Recanati 2012, p. 184). Vertical links are utilised to link regular files with vicarious files, in order to state the common external referent during social interactions. Information can flow from the regular file to the vicarious file when the individual to whom the vicarious file is indexed has the same information.

4.1 Mental files adapted for infant theory of mind: Perner and colleagues' proposal

In recent years, Perner and colleagues (Perner et al. 2016; 2017; Huemer et al. 2017; and see also Rakoczy et al. 2015) attempted to adopt Recanati's theory for putting forward a psychological theory of information storage and retrieval that shows how «the ability to understand belief is a cumulative process» (Perner et al. 2015, p. 87). Perner and coworkers argue that children are able to comprehend to some extent the aspectuality of beliefs only from the age of 4 because of the peculiar formation of their mental file representations. Furthermore, they claim that only from the sixth birthday children are equipped with a full form of aspectuality.

Perner and Leahy (2016, p. 498) state that «children younger than about 4 years can anchor each file to the same object but cannot represent […] the identity of the object referred to by the files». In a few words, infants may lack this horizontal link (enabling to set up the sameness of reference) in a conversation, and even though they are able to represent the same target with different files, toddlers «have no awareness of that» (Perner&Leahy 2016, p. 498). Mental files have the crucial function to encode information from different perspectives that language alone cannot explicitly fix unless the interpreter is helped by contextual cues to individuate the correct referent. Mental files allow individuating the referent in virtue of free information flow among regular

files that result impossible for infants. However, the example of the "bear" reported by Perner and Leahy (2016) concerns linguistic labelling. Therefore, why Perner and Leahy (2016) do use the difficult to manage alternative naming for false belief understanding? Because, according to them, alternative naming and false belief understanding emerge developmentally together. Indeed, for what concerns attribution competence, the "no awareness" state avoiding the individuation of sameness entails also incapacity to establish links between the horizontal level and vertical level before the age of four. In fact, «for a proper understanding of belief one needs the ability to link files that share the same referent» (Perner&Leahy 2016, p. 501). In this case, information can flow from the regular file to the vicarious file when the individual to whom the vicarious file is indexed have the same information. Accordingly, the creation of this kind of link so necessary for attributing mental representations to others requires a conscious reflection which is supposed to be gained by children from the revolutionary passage to the fourth year of life. For each perspective, a new mental file is opened, but for my purpose, it is irrelevant if different mental files about a referent stay actually together. The main focus is turned to the on-line capacity to update, throughout further modes of presentations, the referent features, i.e., the epistemic contents relative to the referent ("something" in the box), which may be clearer along time. There is a transition, possibly described as a *horizontal* or *vertical link*, from a first doxastic state referring to a mental file to a current doxastic state relative to another mental file about the same referent[16]. There is also a distinction between information that fixes the file's referent (*individuating information*) and «information about the referent on the file» (*predicative information*) (Perner et al. 2015, p. 79). Thereby regular mental files can grasp (in Frege's sense) different perspectives on the same object. Perner, Huemer and Leahy (2015) report the example of a butterfly as target object: a mental file is created once the subject is acquainted with the external entity. Thereby, the mental file is anchored to butterfly and stores the predicative information that it flies, is white, etc. Put the case that an object has a dual nature (e.g., being a pen

[16] The idea is simple: two separate mental files exist about a state of the world. One characterises the doxastic state of a subject *before* learning something about the world, for instance, to mention the well known example about the alleged identity between Hesperus and Phosphorus. The other one is referred to the doxastic state of the same subject *after* learning such identity (see Recanati 2012, p. 182).

or a rattle) as it has been proposed in some experiments with toddlers (Fizke et al. 2015; Perner et al. 2015; Rakoczy et al. 2015; Sprung et al. 2007). In this case, we will have different mental files depending on the stay in which the same object is presented.

The «mature mental filing system» can track information available at certain time about an object that may be represented in this way i) from different «*conceptual perspectives*», and ii) from different «*mental perspectives*», i.e. from different others' knowledge about the same object (Perner et al. 2015 p. 79). It is worth noting the use of the term "conceptual" that should pertain also "propositional attitude" that lacks, instead, for the infantile mental filing system. Indeed, as we will see in the next paragraph, a single mental file relative to an object cannot be linked horizontally to another mental file regarding the same object in infant mind. This determines for Perner and colleagues the absence of aspectuality and, thus, of propositional attitudes.

4.2 The crux of aspectuality

According to Perner and colleagues, aspectuality capability enabling to sustain propositional attitude seems to not fully raise at the age of 4. Sprung and colleagues (2007) revisited Apperly and Robinson's (1998; 2003) experiments a developmental window between 4-6 years aged about intentional abilities applied in false belief tasks concerning objects having dual identity (e.g. a die that can be also an eraser)[17]. Sprung, Perner and Michell (2007) predicted that 4-6 year-olds have no problems to pass both the knowledge question ("Does the agent know the dual nature?"), and the standard where-look question ("Where will the agent look for the toy?") only if «information about [the] object was provided in a predicative manner rather than in an individuating manner» (Perner et al. 2015, p. 78). While in Apperly and Robison's (1998) task the

[17] Apperly and Robinson (1998) familiarised 4-6 year-old children with an eraser and an eraser that was also a die. A puppet called Heinz entered the scene, saw the two objects but was not informed about the dual nature of one object. The experimenters asked children whether Heinz knew the eraser was also a die. The children able to pass standard belief test answered "No". However, although they knew that Heinz thought that a die was in location A, and the same time they knew that Heinz ignored completely that the die was an eraser, when the experimenter asked "Where will Heinz look for an eraser?", children chose randomly between location A where the die-eraser was, and location B where the standard eraser was.

child and the agent (puppet Heinz) could both see a die and an eraser but only the child knew the dual nature of the toy - in the revisited task designed by Sprung and colleagues (2007) - the child and the agent could see a long stick, but only the subject knew that the occluded stick was long too. In this case, children were able to answer correctly either the knowledge question, or the where look question.

4.3 Belief attribution for dual identity objects

A simplified version of Apperly and Robinson's (1998) experiments has been proposed by Rakoczy, Bergfeld, Schwarz and Fizke (2015) using a single object with a dual identity (pen/rattle). Children (3-6 year-old) and a puppet called Susi observe together a pen being placed in box 1. Then Susi leaves the scene and in her absence the pen is taken from box 1 by the experimenter who tells children that the pen is also a rattle by showing that it rattles when it is shaken. The object is linguistically termed a "rattle" and it is located inside the box 1 again. Then Susi comes back and witnesses how an unseen object hidden in the experimenter's hands is removed from box 1, called as "rattle", and placed inside the box 2. At this point, children are asked where the puppet will look for the pen. The correct answer is box 1 since Susi does not know actually the dual identity of the rattle. The radical simplification of this test highly correlates it with a standard false belief task. This result denies to some extent the cogency of Apperly and Robinson's experiment, even if it does not explain Sprung et al.'s results. Perner and colleagues' (2015) claim that their hypothesis, grounded on mental file theory, can explain all these data: the deficit of horizontal link can support Sprung and colleagues' data. Instead, for Rakoczy et al. (2015, pp. 496-497) experiment, Perner and coworkers propose that «children ignore links that exist between their regular files when dealing with vicarious files» (Perner&Leahy 2016, p. 505). This is more evident in the true condition where both the children and Susi know the double identity[18]. In this case, if children fail «to take care of links between vicarious files»,

[18] According to Perner and Leahy's mental file analysis, when the child and Susi encounter the pen in the beginning, the child copies her regular file, which represents the pen in box 1, to a vicarious file. Subsequently, «the child learns that the pen is also a rattle, which leads to a regular rattle file anchored to the same object. Finally the puppet witnesses that an object called "rattle", which makes a rattling noise, is invisibly moved from box 1 to box 2. Puppet's *linguistic contact* with the rattle lets children copy their

they will give the wrong answer since they are not yet able to represent that «the puppet knows the identity of pen and rattle by linking puppet's vicarious files. Without linking between vicarious files the verbal information about the transfer of the rattle to box 2 will not be registered on the vicarious pen-file» (Perner&Leahy 2016, p. 506). Sixty per cent of children who passed false belief task answered wrongly that the puppet will look for the pen in box 1.

To sum up, Perner and colleagues (2015; 2016) stress Recanati's mental file theory, and in particular its relational structure, for an adequacy to alleged infant incapacities. The ability to link mental files is the determining factor, and it is *assumed* that only around four years children become able to accomplish it. Without such ability, infants cannot realise that once individuated one way a referent is still the same thing also if it is individuated another way. And at the same time, without the linking ability children cannot establish the connection between a regular file and the vicarious mental files necessary to capture another person's first order and second order false belief, as revealed by the test using dual identity objects. In this view, infants cannot copy regular file to a vicarious file «with the same individuating label» and the predicative contents; furthermore, they cannot catch that «the holder of the vicarious file [should be] aware of the identity» of the referent (Perner et al. 2015, p. 79).

On the contrary, following nativist account, I sustain the aspectuality competence to be early present and at work in infants, in addition to the fact that infants show remarkable flexibility not only for attributing false beliefs about location, but also for attributing other reality-incongruent epistemic states like pretence (e.g., Onishi et al. 2007), false perceptions (Song&Baillargeon 2008), false information about non-obvious properties (Scott et al. 2010), false information about identity (Scott&Baillargeon 2009). In the latter task focused on identity (Scott & Baillargeon 2009), for instance, 18-month-old infants were familiarised with an agent facing a one-piece penguin and a

---

regular rattle file to a vicarious rattle file which registers the move to box 2. Puppet's perceptual contact of hearing the object in the experimenter's cupped hands would also lead to a vicarious copy of the pen file, [but] the already existing vicarious pen file showing the pen in box 1 will inhibit the creation of a duplicate. Since the child understands that the puppet doesn't have information about the object being a pen the outdated information about its location will not be changed and result in a false belief about location» (Perner&Leahy 2016, p. 506).

disassembled two-piece penguin. In each trial, the agent hid a key in the bottom side of the two-piece penguin and then assembled it. The penguin assembled was identical to the one-piece penguin. «In the test trials, while the agent was absent, an experimenter assembled the two-piece penguin, placed it under a transparent cover, and then placed the one-piece penguin under an opaque cover» (Baillargeon et al. 2016, pp. 174-175). Infants managed to attribute to the agent two interlocking false beliefs.

However, to give much strength to my argumentation, I need to present analogue experiments based on a supposed awareness of dual identity object for children much younger than 3-5-year-old. For this purpose, I am going to illustrate in the next two paragraphs another attempt advanced recently by Fizke and colleagues (2017) to verify the shortage of aspectuality skills in 2 year-olds. Against this view, I will present Buttelmann et al. (2015) helping task aiming «to measure infants' understanding of others' belief about the identity of an object» (Buttelmann et al. 2015, p. 96). Finally, an experiment based on tools very similar to Rakoczy et al. (2015) and Perner et al. (2015) studies, and designed by Kampis (2017) for testing whether 14 month-olds are able to understand aspectuality of belief representations.

4.4 Aspectuality of belief representations in two helping tasks

Fizke and colleagues (2017) designed an experiment with 26 month-olds following the active helping paradigm sketched by Buttelmann et al. (2009). They reproduced two conditions: a standard change-location condition using two different toys and a special box with several openings, and an aspectual-condition in which they used one toy, a bunny that could be transformed into a carrot. In each condition they created a true belief and a false belief context, in order to compare the eventual different toddler's behaviour. In non-aspectual change-location condition an experimenter (called E2 in the original paper) expressed liking for two toys that she found on a table next the box, then she puts them inside the box because she announced the need to leave. At this point, in the false belief situation when she left, another experimenter (E1) appeared behind some curtains greeting the child, and she proposed playing a jest on E2. «To this end, E1 took one of the toys out of the box and sneakily hid it under a tissue» (Fizke et al. 2017, p. 218), located on the same table. In the true belief situation, E2 did not leave the room, and E1 invited E2 to attend her action. At the end, E2 told the child that she

had to leave, and left the room. «On her return (in both conditions), E2 approached the table and reached into the box. She took out the remaining object and put it beside the box. She then began to search in the box again, saying, ‘‘Hmm, eh? I don’t understand ... but where is ....”» (p. 219). If the child did not react, E2 exaggerated her disappointment, saying: ‘‘Hmm. Oh no!”. At this point, if the child did not react yet, E2 asked: ‘‘Can you help me?”. E1 could also elicit and suggest child helping E2. Fizke and co-operators confirmed Buttelmann et al. (2009, p. 339) study since the most of young children helped E2 to open and search into the box in true belief context, believing that maybe E2 had some goals to reach again into the box even if she knew that the other toy was under the tissue. Whereas in false belief condition, most of infants pointed or gave the toy they believed E2 was searching for into the box.

Instead, in the aspectual true and false belief condition, the researchers used only one toy but with twofold identity. First of all, E1 showed how the bunny could transform into a carrot, saying: «‘‘Look! The bunny is also a carrot!” In false belief condition, she did this in a sneaky way, whispering ‘‘Shh” and telling the child, ‘‘[E2] does not know that, right?”. The child was then asked ‘‘Can you make it so that it is a bunny again?” (and was helped by E1 if the child was unwilling or unable). E1 then placed the toy in its original aspect (e.g., bunny) on the table and returned to her place behind the curtains. After this, E2 came back, found the toy, expressed her liking for it, and put it into the box before leaving again» (Fizke et al. 2017, p. 219). In true belief condition, E2 attended before leaving how E1 transformed the bunny into carrot (or vice versa). In either the absence (Aspectual FB) or the presence (Aspectual TB) of E2, E1 returned and took the bunny (for example) out of the box, transformed it into the carrot, so she put it back into the box. When E2 entered the room for the second time, she reached into the box, took out the carrot putting it, and finally she began to search again. The test trial was identical to non-aspectual conditions, but this time there were not relevant behavioural differences between true and false belief condition: some infants helped to search in the box, and others gave E2 the toy.

According to the authors, these results may indicate the infants’ lack of «ascribing fully fledged propositional attitudes». It seems that infants have not been able to distinguish between true and false condition. However, I deem that in this case the active helping task may be misleading for the comprehension of the goal that infants

should be inferred. Buttelmann et al. (2009) study relies on the understanding of searching for action's goal both in true and false belief condition[19]. Such understanding, that refers to the actual agent's goal, might be overwhelmed by the more complex computation of the putative false belief inference about the toy identity, and its false location. Moreover, in light of Helming et al. (2016) study, the fact that toddlers receive the burden of the verbal helping direct request may constitute a further challenging cognitive process. Given all these factors, infants might think that both solutions are correct, or in other terms, both solutions represent the correct way to help the agent achieve her scope. Therefore, Fizke and colleagues' findings do not reveal clearly infants' incapacity to sustain sophisticated belief reasoning.

A further proof of infant false belief flexibility in the regard of dual object identity recognition comes from another helping task experiment conducted by Francis and David Buttelmann with Janina Suhrke (2015), and «designed to measure infants' understanding of others' belief about the identity of an object» (Buttelmann et al. 2015, p. 96). Specifically, they tested whether 18 month-olds may have simultaneously a double representation of the apparent and the real identity of an object and whether they enable to attribute these representations to another person. In their experiments, they used four deceptive objects: a sponge that looked like a rock, a box that looked like a book, a pencil which looked like a branch, and a brush that looked like a duck. It is worth noting that these objects could be considered as toys; the duck, for example, was also really a brush, i.e. it could function as a brush. Therefore, in my opinion, there is no significant difference with the bunny toy that can be also a carrot used by Finzke and colleagues (2017). All these objects are represented as toys with twofold aspects and functions. For each of the deceptive objects, there were the real objects, a duck toy and a brush, a pencil and a branch and so on. First of all, infants familiarized with the hand-giving gesture of the experimenter to prime helping behaviour both in presence of the

[19] Eighteen-month-olds are instructed to lock and unlock two boxes. A male agent enters the room and hides a toy in a box, then he leaves. In the agent's absence, the experimenter transfers the toy to the other box, and locks both boxes. When the agent returns, he tries unsuccessfully to open the box where he put the toy. Most of infants approach the other box for helping the agent to achieve his goal (i.e. to retrieve the toy). This may indicate that infants comprehend the goal and realise that the agent falsely believes the toy is still in the prior location.

experimenter (E) and of an assistant (A) who asked infants: “Can you give this to [E]?” (Buttelmann et al. 2015, p. 98). After familiarization, A showed a deceptive object both to infant and E, who manifested great interest in it. Then, in true belief condition, A explained the «unexpected real identity of the deceptive object», while in false belief condition «E remained outside while A demonstrated the unexpected real identity of the deceptive object». At that moment, while A put the object onto a shelf, E reentered the room manifesting the desire to have something. She started to reach the deceptive object on the shelf but unsuccessfully, so she asked A to help her. A pretended to be involved in other things, and E expressed her disappointment and stopped reaching the object. At that point, lifting up an occluder, i.e. a cardboard from the floor in front of the shelf, A revealed two objects: «one resembling the apparent identity of the deceptive object and the other resembling the real identity (i.e., the function) of the deceptive object» (Buttelmann et al. 2015, p. 99). Then, A asked the child to help E by getting what she desired. Infants had to infer the E’s goal, and to perform the correct helping act, they needed to hold E’s belief about the object’s identity. In false belief situation, E ignored the mismatch, therefore she wanted to reach the object with deceptive features (e.g., a stone, or a duck). On the contrary, in true belief condition, infants had to suppose that E wanted the object that corresponded to the real identity (e.g., the sponge and not the rock).

The results reflected these expectations because they differed remarkably between the two conditions. In fact, in false belief condition, infants gave E the object that looked like the apparent form (e.g. the rock and not the sponge) more often (64,6% of trials) than in true condition (34% of trials). Infants understood that «E believed the object to be what it appeared to be [e.g. a branch, or a rock, or a duck], and brought the object that resembled the appearance of the deceptive object» (Buttelmann et al. 2015, p. 100). Whereas in true condition, infants understood that E knew the real identity of the object, and since they identified E’s end, they gave E «the object resembling the real identity significantly more often than the object resembling the appearance» (p. 100).

This finding shows how infants are able to use person’s belief about the object identity in order to infer her goal and to help her accordingly. To accomplish this complex target, infants have first to distinguish the double identity of the objects. Therefore, according to this research, children younger than 4 years of age demonstrate

to possess abilities belonging to System 2 (following the terminology of e.g., Apperly&Butterfill 2009, Fizke et al. 2017), because, in the case of the prior experiment, they understand how the experimenter represents the target object depending on the different conditions. In conclusion, such experimental results suggest that the early understanding of others' false belief can be fitted across a variety of situations and tasks as in the following experiment that I am going to report briefly.

4.5 Kampis' (2017) test of false belief about object identity[20]

Researchers presented 14 month-olds «dual-identity objects that could transform between two appearances» (Kampis 2017, p. 200). The aim was to test whether 14-month-olds «who can selectively vary the use of appearance information […] can also attribute to another person a false belief about object identity, when this belief is based on mistaken individuation» (Kampis 2017, p. 186). They used for belief trials toys with twofold forms, a frog that could transform in crab, a bird in a hedgehog; while for the baseline condition two other transforming objects were used differing in colours and materials (a princess in a duke, and a fairy in a prince, but without resembling humans or any particular animal). The baseline trials consisted of two conditions: *unknown transform*, in which infants could not see that the experimenter transformed by her hands the princess (for instance) in the duke just inside the opaque box; known-transform condition in which experimenter showed the dual feature of the toy. Thereby, through the baseline Kampis figured out which infants understood that:

> «(i) if two objects seem to be of different kinds, then they are likely two different objects, and (ii) if there is evidence that one object can appear in two different forms, then this overwrites the inference from the first point and there is likely only one object. [Indeed] in the unknown-transform trials infants should infer to be two objects (and hence one remaining at the time of search), whereas in the known transform trials infants should come to the conclusion that the two forms belonged to the same object (and therefore none remains by the time of search)» (Kampis 2017, p. 194).

[20] I report such intriguing experiment from Kampis' doctoral thesis (2017, pp. 186 and followers).

Kampis assumed that those infants were able to compute the dual identity of the object, then they would also successfully represent the equivalent scenario from someone else's perspective. She measured the belief attribution indirectly, «through assessing whether infants' search duration in the box varied depending on the other person's belief about the content of the box» (Kampis 2017, p. 200). In test trials an experimenter (henceforth E2) did take out an object from her bag, put it on the top of the box, and pointing at it she said: "Look, a bird!". Then another experimenter (henceforth E1) repeated the label: "Oh, a bird!", and put it into the box. Then, in false belief trial, E1 said: "Oh, my phone is ringing, I have to run out", and so left the room. During E1's absence, E2 said to the infant: "Look!", and, reaching into the box, she retrieved the toy for demonstrating the transformation into hedgehog ("Let me show you something! The bird [or the frog] is also a hedgehog [a crab]! Do you see?"). After the transformation, she located the toy into the box again. When E1 came back, she searched for the toy into the box (as in the familiarization trial). Once the toy retrieved, she said: "Oh, a hedgehog [a crab]! How nice!"; so she put the toy away into a bag, and took a book saying: "I have to look up something now", and pretended to read for 15 seconds. After the 15s E1 said: "Ok, we are done". Then, E1 gave the toy to E2 to put it away, but before putting it away E2 showed it to E1 and said: "Look, let me show you something. Do you see the [crab]? The [crab] is also a [hedgehog]"» (Kampis 2017, p. 192). Crucially, infants were allowed to search inside the box at this time, so they could have evidence that the hidden objects have been retrieved, and necessarily there was nothing in the box. Kampis measured the time infants spent in this search, predicting – in line with Kovács et al. (2010) - that infants would have spent more time when the other person had a false belief about the content of the box. Indeed, in the false belief test trials «the other person believed that one of the objects was still in the box, as she had not seen the object transform […]; whereas in the true belief trials the other person knew that there is no object in the box» (Kampis 2017, pp. 201-202). Infants searched longer in false belief condition, than when the other person saw the transformation and knew the content. This modulation of infants' own behaviour, already predicted by prior Kovács and colleagues' study (2010), may constitute a further proof that «infants' own representations and the ones they attribute to others have a common representational format»; and furthermore we can claim that without «metarepresenting the other

person's representation of two objects», infants would not have succeeded (Kampis 2017, p. 205).

***

The acceptance of representational cognitive structure of infant's mind does not imply the solution of the developmental puzzle whereby infants do not manage to pass successfully explicit false belief tasks. The most interesting and decisive proposal has been proposed by Helming and colleagues' (2014; 2016) account that involves many elements featuring natural pedagogy system, but before introducing their hypothesis, it is worth presenting a neuroscience perspective about the interpretation of this enigma.

5. An unitary TOM observed from a neuroscience perspective

> «Much remains to be discovered about how infants' ability to infer and reason about others' mental states improves with age, about the maturation of the brain networks that underlie this ability, and about the various factors that contribute to individual differences in neurotypical and other populations» (Baillargeon et al. 2016, p. 179).

One of the certainties that appears to be gained thanks to the ongoing neuroscience investigations is that the implicit-explicit TOM distinction loses, or at least fades, its neatness if we investigate brain network. As Baillargeon and her colleagues (2010, p. 115) have put it, the spontaneous-response tasks request false-belief-representation process without other kinds of cognitive efforts such as response-selection, and response-inhibition processes that could overwhelm the infant limited resources (for further details see Carruthers 2013). Neuroscience findings have indicated a possible explanation that relies on an immature state of neural connections between frontal and temporal brain regions in human infants (Johnson 2001; Lebel et al. 2008[21]). Indeed, according to several researches (e.g., Kobayashi et al. 2007; Sabbagh et

[21] Lebel and colleagues (2007, p. 1044) used a brain imaging technique particularly sensitive to axonal packing and myelination. In their large age-distributed sample (202 subjects from 5 to 30 years), through the analysis of white matter changes, they could measure remarkable brain regional variation, and in

al. 2009; Saxe&Wexler 2005; Sommer et al. 2007[22]) the right temporo-parietal junction (rTPJ) plays a crucial role in adults' and children's false belief representations process[23], while the anterior cingulate cortex (aCC) and the medial prefrontal cortex (MPFC) play an important role in the response selection process. The neurodevelopmental delay of the connections between this kind of functional brain areas (for a review of such interpretation see Scott et al. 2010) might generate the insufficient capacities to make the right inferences.

Recent ERP studies (Van Overwalle&Vanderkerchve 2013) conducted with adults have revealed that typical inferences «triggered by implicit and explicit instructions have a similar early timing» (Kampis et al. 2017, p. 55). Van Overwalle and colleagues conducted also functional imaging studies to investigate the overlap between explicit and implicit mentalising, and they found a common underlying mentalising network «that is relatively blind to the implicit or explicit nature of the inference, and that seems more sensitive to the content of the inference» (Van Overwalle&Vanderkerchove 2013, p. 2). Bardi and her colleagues (2017), analysing «the BOLD signal for false belief processing by directly comparing spontaneous and explicit task versions», found that neural mechanisms involving TPJ and anterior MPFC areas overlap both in spontaneous and explicit TOM (Bardi et al. 2017, p. 391). Another fMRI study (Kovács et al. 2014) in accordance with a research based on NIRS technique (Hyde et al. 2015) confirm that the TPJ (normally involved in explicit tasks)

---

particular, they observed differences in developmental timing that suggest «a pattern of maturation in which areas with fronto-temporal connections develop more slowly than other regions».

[22] In particular, Sommer and colleagues (2007) conducted a fMRI study on sixteen adults (mean age 26) comparing false belief reasoning with true belief reasoning in parallel tasks. They used the typical Sally-Ann scenario through a variety of cartoon stories. «The false belief versus true belief contrast revealed activation of the dorsal part of the anterior cingulate cortex (dACC), the right lateral rostral prefrontal cortex and the right [temporo-parietal junction] (RTPJ) associated with false belief». They suggested that the dACC and the lateral prefrontal cortex (PFC) might be involved with «action monitoring and stimulus-independent cognitive processing whereas the activation of the TPJ might be related to the computation of mental representations that create perspective differences» (Sommer et. al. 2007, p. 1378).

[23] Biervoje and colleagues (2016, p. 1) reported the «causal link between a specific subregion of the TPJ and a specific cognitive facet of ToM» through the analysis of «two brain-damaged patients whose common lesions were almost exclusively in the left posterior temporoparietal junction (TPJp) and who both showed the same striking and distinctive theory of mind (ToM) deficit».

is also triggered during implicit belief processing, but with a remarkable difference emphasised by Kovács and colleagues (2014): only events involving agent's false beliefs about the presence of an object elicited activation in the TPJr (typically committed in explicit false belief test), and not agent's false beliefs about the absence of object. This evidence leaded the researchers to conclude that:

> «While humans can explicitly attribute to a conspecific any possible belief they themselves can entertain, implicit belief tracking seems to be restricted to beliefs with specific contents, a content selectivity that may reflect a crucial functional characteristic and *signature property of implicit belief attribution*» (Kovács et al. 2014, p. 1).

A recent and revolutionary fMRI study conducted by Richardson and colleagues (2018) on 3-year-olds arrives at the same conclusion: passing explicit false-belief tasks in a later developmental stage «does not correspond to discontinuities in the neural basis for reasoning about the minds of others» (Richardson et al. 2018, p. 8). For the first time, it has been possible to analyse a large sample (n= 122) of children between the ages of 3–12 years old compared with a reference group of thirty-three adults. Every one of them watched a short, animated movie that did not require learning a task, «and included events evoking the mental states and physical sensations of the characters, while undergoing fMRI» (*Ibid.*, p. 2). The researchers compared TOM network with pain[24] brain network (bilateral medial frontal gyrus, insula, and secondary sensory cortex, and dorsal anterior middle cingulate cortex[25]), and «report evidence that ToM and pain networks are functionally distinct by 3 years of age, and become increasingly specialized between the ages of 3–12 years» (*Ibidem*). Moreover, they highlighted that:

> «Brain regions involved in ToM in adulthood *already constitute a distinct network in 3-year-old children*, which gradually becomes more integrated and distinct from other networks over the next decade. […] Focusing specifically on 3- to 5-year-old children, *the neural responses* to social movies in children who systematically fail

[24] Pain is intended as the perception of physical pain and bodily sensations of others (Richardson et al. 2018, p. 2).

[25] See Zaki et al. 2016.

versus pass explicit false-belief tasks *were similar*» [my italics] (Richardson et al. 2018, p. 8).

We can conclude that neuroscience perspective is much more compatible with a continuum development of TOM hypothesis supported by the growth of neural connections among specific areas, rather than with the existence of two-systems, or the incapacity to mobilise sophisticated cognitive resources

6. Attempts to solve the developmental gap from a cognitive point of view: the referential communicative bias

«Why do most children fail standard elicited-response false-belief tasks until they are at least 4 years old, while the looking behavior of preverbal infants strongly suggest that they can track the contents of others' false beliefs about an object's location?» (Helming et al. 2016, p. 438).

This is the developmental paradox emerged by the comparison between implicit and explicit false belief test. From a cognitive point of view, all the experiments presented by nativist theorists do not provide a direct solution to developmental discrepancy puzzle whereby children across countries fail at Sally-Ann standard false belief tasks until age 4 (e.g., Callaghan et al. 2005; Liu et al. 2008; Sabbagh et al. 2006; Wellman et al. 2001).

Helming and colleagues (2014; 2016) suggested that the developmental puzzle can be overcome by avoiding a direct asking attitude. Indeed, asking the question directly "Where will Sally look for her toy?", the experimenter is referring to the toy «while she shares the children's correct perspective on its actual location» (Helming et al. 2014, p. 169). The researchers hypothesised that the prediction question posed by the experimenter might be interpreted by the child as a normative request on the basis of what the child knows about the actual scenario, as if it were: "Where should Sally look for her toy?" (Helming et al. 2014, p. 169). Therefore, in order to pass a verbal explicit design task, it is enough to address young children verbally in different ways as suggested by Scott et al. (2012), and Rubio-Fernández&Geurts (2013). For instance, Rubio-Fernández and Geurts (2013, p. 31) designed a new verbal false-belief task that

allowed to keep «track of the protagonist's perspective during the false-belief narrative». Their experiment involved a puppet (called *Duplo girl*), who has a false belief about the location of her bananas. Three-year-olds, who know the actual location of the bananas, are prompted to act out the puppet's most likely action by the experimenter who tells them: 'What happens next? You can take the girl yourself if you want. What is she going to do now?'[26] In response, the majority of 3-year-olds move the girl to the empty location in accordance with the content of her false belief (for a review of the experiment see Helming et al. 2016, p. 445). Children had not to choose between two possible answers, rather they were asked open questions that invited them «to continue acting out the story» (Rubio-Fernández&Geurts 2013, p. 31). The researchers achieved the striking result that 80% of young children with a mean age of 3.5 years were able to pass their verbal task.

Scott and colleagues (2012) tested 2.5-year-old toddlers using a preferential-looking task and a violation-of-expectation task. In the first task toddlers «listened to a false-belief story while looking at a picture book (with matching and non-matching pictures)», while in the second task «children watched an adult 'Subject' answer (correctly or incorrectly) a standard false-belief question» about the location of Sally's toy ("Where will Sally think her toy is?") (Scott et al. 2012, p. 181). The obtained results showed that children understood both false belief scenarios despite the linguistic demands which were not addressed directly. Keeping a third person perspective on the event, toddlers showed to have understood when Subject held a false belief thereby passing successfully verbal false-belief task without answering direct questions about agent's beliefs.

6.1 Helming, Strickland and Jacob's solution committed with the natural pedagogy theory

In 2006, Csibra and Southgate already suggested that the 'where' question addressed in the standard verbal false belief test could be interpreted by infants as referring to the final hidden object location, rather than to the expected agent's action (Csibra&Southgate 2006). This suggestion has been employed by Helming, Strickland

[26] It is to be noticed that no belief-verb was used. This caveat is important for the next paragraph.

and Jacob from 2014 and subsequently enriched by the three authors in 2016. Helming, Strickland and Jacob (2014) tried to provide a solution to the puzzle about early belief ascription articulating a *pragmatic framework*. According to them, young children do not pass elicited-response false-belief tasks, despite the fact they show spontaneous false-belief understanding, because even if infants comprehend others' false beliefs, they are overwhelmed by tackling two distinct action interpretation processes: «the instrumental action of a mistaken agent and the experimenter's communicative action» (Helming et al. 2014, p. 167; 2017). Such overwhelming would generate in young children a «reality bias» triggered by the cooperative communicative interaction. In their view, failure in an elicited-response false-belief task does not mean for the young children the incapacity to represent the contents of others' false beliefs. The failure is due to the cognitive implications for a young child that a communicative relation in second person view entails. Specifically, according to Helming and colleagues (2016, p. 449), «the systematic failure of most children under 4 years of age in elicited-response change-of-location false-belief tasks» do not reflect «their inability to track the contents of others' false beliefs», but rather it is due to a tension between the second- and the third-person perspectives. In other terms, «young children are required to take a second-person perspective on the experimenter's communicative action» (Hemling et al. 2016, p. 453).

Therefore, the young children ability to maintain the third person-perspective and to track others' false belief breaks down when the shift from third to second-person view occurs. This is the case predicted by natural pedagogy theory, and indeed Helming and colleagues refer to the Csibra and Gergely's theory as a potential explanation for such cognitive biases which avoid children to pass the standard verbal false belief task. Such intriguing suggestion highlights the conflict between mindreading and natural pedagogy systems. In the third chapter, we have seen how a triadic relation founded on a direct (and ostensive) communication triggers special biases on infant's mind that guide and affect infants to interpret the subsequent agent's instrumental action (like for example lighting the "magic" box with the forehead rather than with the hands). Therefore, when infants are involved in elicited-response false belief tasks, «they are required to stick to a third-person perspective on the mistaken agent's instrumental action» (Hemling et al. 2017, p. 453). This provokes a *cognitive tension* whereby a

young child's mind generates two biases: one «highlights the epistemic perspective on the object's actual location that is shared by the children and the experimenter, at the expense of the mistaken agent's different epistemic perspective on the object's location. The other bias motivates young children to help the mistaken agent achieve the goal of her instrumental action» (Ibidem). The former is termed *referential bias* triggered by the presence of an ostensive signal that the recipient interprets as directly addressed to her by the communicative agent. This could induce the infant addressee to assume a pedagogical attitude whereby she predisposes to receive epistemic contents by the communicative agent.

According to Hemling and colleagues (2016, p. 456), there are two competitive epistemic perspectives that young participants to standard false belief test have to take into account with the respect to the reference of the location of Sally's marble: Sally's mistaken perspective, or «the epistemic perspective shared by participants and experimenter with whose communicative action they are currently engaged from a second-person perspective». The latter induces to infer the marble's actual location, the former leads to the empty location. The pedagogical condition created in virtue of the communicative relation would entail that the shared epistemic perspective trumps Sally's mistaken perspective.

Furthermore, the primarily epistemic cognitive effect triggered by such pedagogical condition, i.e. the direct communicative relation, determines the onset of a normative bias whereby the infants might assume that the experimenter is asking them to report where the mistaken agent *should* look for the marble, and not where she thinks to be (Hemling et al. 2016, p. 458). Infants may have formed a false belief representation ascribed to the agent, but they do not use it to answer the experimenter because the question is differently interpreted given the communicative relation between the experimenter and the infants relatively to an agent's action. The misinterpretation does not depend on lack of linguistic capacities, or metarepresentational competences. The Southgate et al. (2010b) study may help us to highlight this aspect.

6.2 Grasping the relevant communicative referent

Seventeen month-olds familiarised themselves with two previously unfamiliar objects, one of this named "sefo" (a nonsense name) by the experimenter. The two toys were placed into two distinct boxes, so the experimenter left the room. At this point, another experimenter entered the room and changed the objects locations. Then, the first experimenter returned, telling the infant she likes *sefo* very much and wants to play with it, so pointing at the box in which the experimenter believes the *sefo* was, she asked infant: "Can you pass me the sefo?", or "Can you pass me it?". Most of the infants moved toward the not indicated box, interpreting correctly the experimenter's referential intention in the light of her false belief about the actual *sefo*'s location. This result shows that infant identified and recognised the object as a specific kind of object distinct from another, and understood that the agent held a false belief about which object was in which box (Southgate et al. 2010b; see also Carruthers 2013, pp. 145-146). Furthermore, these results indicate that before the end of their second year, children are able to track the content of an agent's false belief and to use such representation for helping the agent achieve her goal from a second-person perspective (Hemling et al. 2016, p. 452). But in this case, the kind of links between experimenter, infant and reference is different: the ostensive communication does not establish a pedagogical stance, potentially guaranteed to some extent in standard explicit false belief test, between the experimenter and the infant, rather Southgate et al. study could be described as a false-belief helping task, where the instrumental action is gained through direct asking. In this way, the crucial factor for the infant's failure is not the verbal request as such, rather the modality of the direct request in line with Rubio-Fernández&Guertz (2013) and Scott et al. (2012).

6.3 Westra and Carruthers' proposal: from referential bias to semantic incompetence

Westra and Currethers (2017) provide a very similar pragmatic explanation too, based on a pedagogic intentional exhibition. When the experimenter refers to the target object she is drawing children's attention to its actual location, inviting them either to

exhibit their knowledge about the object actual location, or to help the protagonist rather than revealing their knowledge about the protagonist's belief. The divergence with Helming and colleagues' model regards the fact that according to Westra and Carruthers (2017, p. 169), «the child continuously represents the agent's false belief throughout the task» and the reference bias does not interfere with mindreading process triggered in third-person view. Simply, infants do not use this kind of information to answer experimenter's question since they have the other two salient interpretations available. For the novice speakers, linguistic interactions about others' belief are novel and communicated by the use of verbs as "to think" that may be intended as "to want". In other words, infants might have difficulties to understand the real topic of talk that is deviated elsewhere, likely towards the world and not the subject's mental state. For example, sentences with the form "S thinks that P" (e.g., "Scott thinks his mittens are in the closet"[27]) might draw their attention to ''P" ("mittens are in the closet") rather than making them attribute a belief to the subject, a false belief in this case because the mittens are in the backpack (Westra&Carruthers 2017, p. 170).

On closer inspection, semantic incompetence interpretation does not exclude the referential bias model, if we consider Scott et al. experiment (2012). Both models could be adequate according to the ages of children, the presented scenario and the contextual verbal interactions. Thanks to Kampis et al. (2013; 2016, Kampis 2017) and Kovács (2016), we have emphasised the precocious stance to keep separate belief content from belief holder, and the onset of language comprehension does not seem to attenuate this tendency immediately. Rather, what we can figure out is that the development of pragmatic skills allows discerning when and with which terms cognitive states become the topic of conversations. Contextual training improves the exact interpretation of such verbal exchanges possibly uttered with direct and indirect speech. So infants have to learn that sometimes questions are really invitations for them to display their psychological knowledge rather than requests to be helpful or to display their knowledge of the worldly facts[28].

[27] Sentenced used by Wellman and Liu (2004) in their TOM scale tasks, and quoted by Westra&Carruthers (2017, p. 170).

[28] An Italian research of University of Pavia conducted by Lecce and coworkers (2014) showed that after particular intervention using TOM tasks, 4-5 year-old children improved their first-order false belief

7. Conclusions

In conclusion, infants can hold other's mental states through representational formats, whose contents involve location, identity and absence of objects that ceased to exist. Representing propositional attitudes could be easier for infants than entertaining other forms of representation, like registrations (Carruthers 2015). The format of mental state representations presents a syntax as propositional attitude composed by an agent, an attitude, and a content. These components can be changed flexibly in early TOM operations as suggested by some evidence (Kampis et al. 2013; Kampis 2017) and described by belief file theory (Kovács 2016). The two variables of belief file (the agent and the content) «can be modified independently from each other, which enables fast updating and modification of the elements» (Kampis 2017, p. 66), allowing fast computation on any belief content. On the contrary, «the limit of the speed and effort of computing a particular mental state content would depend on the difficulty to calculate and represent such content itself» (Ibidem). Furthermore, Kampis (2017) suggests that also *attitude* can change in the sense that prediction of an agent's action rests on a preference for an object, accompanied by *a belief about* the same object.

Kovács' belief file structure fits well with Carruthers' (2013) description of TOM as a representational multi-componential process that operates through interactions between a core attribution system and the planning system. Accordingly, the attitude component «must be distinct and detachable from the content of the attitude», and only the latter become «available to the planning system, either as a goal to be achieved or as an assumption to be relied upon» (Caruthers 2013, p. 148). The cooperation and the enrichment of other systems make mindreading more and more sophisticated throughout childhood. The innate starting point of TOM's maturation would be, thus, the capacity to represent propositional attitudes. This is in line with Scholl and Leslie's account (1999), in particular when they claim that «TOM has a specific innate basis» (Scholl&Leslie 1999, p. 134), which means that TOM processing

---

understanding significantly more than children in the control condition with no intervention. Crucially, «the positive effect of the TOM intervention was stable over 2 months and generalized to more complex ToM tasks» (Lecce et al. 2014, p. 2404).

does not apply to other cognitive domains, and that TOM «is given as part of our genetic endowment» (p. 134). For this reason – Scholl and Leslie explained – we can define early TOM as *modular* without implicating neither that «the entirety of TOM is modular», nor that «modularity is defined only in terms of restrictions on informational access, and there is no requirement that the process "inside" the module do not develop» (p. 136). In fact, one of the main characteristic of a modular cognitive mechanism rests on its *universality across individuals* (at this regard see also Callaghan et al 2005). The normal development of TOM capacity respects this assumption, even though with differences and delays for each developmental stage: environmental interactions affect the time-table «with which the modular capacity manifest itself» (Scholl&Leslie 1999, p. 136).

Finally, it is reasonable to suppose the existence of a single mindreading system operating, in simple cases automatically, in others spontaneously as suggested by Carruthers, and including propositional formats, attribution procedures and prediction strategies.

> «On the nativist account, children's cognitive development enriches their mindreading abilities: as their own belief forming mechanism matures and they themselves acquire novel beliefs about richer and more complex subject matters (in particular, via verbal communication with knowledgeable speakers), mindreaders also become able to ascribe to others new beliefs with richer and more complex contents. But the basic mindreading system at work is one and the same» (Jacob 2015, p. 3)

One-system account predicts that surely infants lack the (adult) conception of belief that is enriched through learning. However, the notion of *aspectuality* seems to not be the discriminant of this developmental discrepancy, rather a *metacognitive reflection*, i.e. an introspective self-awareness, could mark much better the difference. In other words, the developmental change does not involve representational capacities as such; thereby there is no a real remarkable representational shift, rather *an enrichment of the contents that are represented*. The aspectuality is processed spontaneously without involving an introspection machinery turned toward the content of metarepresentations. The infantile dimension of belief houses not only *aspectuality* but

also *normativity*, in the sense that attribution procedure follows the transitional inferential rule that someone *should* believe something given *that* evidence.

Some fundamental features of belief states are, thus, shared with the adult notion of belief, without implying a well formed and mature mindreading system at birth. Rather it entails, as the nativist account predicts, a developmental system endowed with some characteristics and functions like the extreme flexibility, which allows the rapid construction of a knowledge common ground in virtue of a belief attribution mechanism. The latter funds also the universality assumption committed with natural pedagogy system. Therefore, the early form of mindreading serves social learning practises and it is closely connected and sometimes contrasting with natural pedagogy whose grounding elements seem to be involved in several communicative contexts.

CHAPTER 6.

Limits of natural pedagogy model

1. Introduction: epistemological side of natural pedagogy

Human environment is characterised by informational cooperation and communication, so the advantages of using others' information are extremely high, but equally, the risks to be deceived are too great to be overlooked or underestimated. *Natural pedagogy is based on trust in others*, and this fact deserves further reflections. In the present chapter, my intent is to focus on the epistemological implication of such dependency. I suggest that the information transmitted pedagogically has to be treated as testimony for the reasons that Harris and Lane (2014, p. 444) nicely summarised in the following passage, in which they underline the conditions of children's learning where testimony properly occurs:

> children's learning from the testimony of other people will typically occur when two conditions are fulfilled: (i) the entity or phenomenon is *difficult for children to observe autonomously* or is ambiguous so that, in the absence of definitive perceptual input, they are *receptive to the testimony of other people*; (ii) children can nonetheless imagine—mentally represent—the state of affairs that is described via the testimony [my italics] (Harris&Lane 2014, p. 444)

From an epistemology of testimony perspective, I would like to emphasise the fact that infants are *dependent on* and spontaneously *look for* reliable sources of information and, for this purpose, are equipped with some sort of selective vigilance both toward the informants and toward the contents of message.

I hold that the boundaries of application and the efficacy of natural pedagogy account can be more evident if they are examined under this light. Simultaneously, it will be easier to define the limits of natural pedagogy system's work throughout the development. In fact, when children become sufficiently able to judge the reliability of an informant and the trustworthiness of a message in virtue of inferential capacities and quite robust baggage of beliefs, then natural pedagogy will be inhibited to trigger its

characteristic biases and will be cease to operate. Definitively, *we can define natural pedagogy model as a developmental temporary form of cross-generational transmission of testimony*, where testimony represents – psychologically speaking - «a source of basic beliefs»[1]. Therefore, natural pedagogy system allows *the formation of a stable cluster of cultural shared beliefs*. In this regard, Peter Graham emphasises the role of testimony as that «process of forming beliefs on the basis of understanding what other people say» (Graham 2006, p. 105). As we know, human beings do not limit to entertain beliefs. We assess beliefs and interconnect them inferentially (e.g., Mercier&Sperber 2011; Sperber et al. 2010; Sterelny 2012).

> «Once we have a folk logic installed, we can assess messages for their coherence with what we know from other sources, and with what the agent has previously said. We can build epistemic profiles of agents: *we assess the reliability of sources, as well as the plausibility of messages*» [my italics] (Sterelny 2012, Chapt. 6.1)

In a very recent paper, Pesch, Suárez and Koenig (2018, p. 38) have underlined «the ubiquity of testimony in social interactions», where testimony is intended as «a promiscuous source of knowledge» without which «we could not participate in our shared cultural and social practises». The epistemologist Robert Audi emphasises that: «testimony is a pervasive and indispensable source of knowledge» (Audi 2015, p. 217), and «natural source of beliefs» (Audi 2015, p. 235).

In the subsequent paragraphs we will follow this double rout suggested by Sterelny, namely the *reliability* of informant and the *plausibility* (or *trustworthiness*) of transmitted messages. First, it is necessary to point out the crucial aspect of belief foundation through testimony.

[1] Audi 2015, p. 230.

1.1 Knowledge without justification. «The autonomy of knowledge with respect to justification»[2]

Two principles are applicable to testimony. The first principle concerns *knowledge*, whereby a belief that *p* based on testimony provides that «the attester knows that *p* and the believer has no reason to doubt either *p* or the attester's credibility concerning it» (Audi 2015, p. 227). This is the case of «*testimonially based knowledge*», that implies in turn, on the one hand, that the «grounding of true beliefs is transmissible across testimony» (Audi 2015, p. 227); on the other hand, that «everything known (even in part) on the basis of testimony must [not necessary] be known by someone entirely on another basis» (Audi 2015, p. 228). The second principle regards *justification*, whereby a belief based on testimony provides «the believer has overall justification for taking the attester to be credible» about *p* (Ibidem). In order to provide robustness to these principles and, thus, to testimonial knowledge, the latter *must* ultimately depend on non-testimonial knowledge. In other terms, «the attester's knowledge […] cannot ultimately be grounded wholly in testimony» (Audi 2015, p. 228). Testimonially based knowledge depends (epistemically) on «the potential cooperation of another source of knowledge». Such cooperation is an act of trust, whereby the recipient relies on the reliability of the attester.

> I cannot acquire justification for believing something on the basis of testimony unless I have some degree of justification for believing that the attester is credible (Audi 2015, p. 229).

In other words, testimony as such «does not produce justification in the recipient by transmission» (Audi 2015, p. 249). Indeed, the credibility of the subject, who attests that *p*, is what justifies recipient's believing *p* on the basis of her testimony. Testimony begins with the expression of the attester's knowledge. The expression is, we can say, the *transmission*. In this sense, Audi claims that «testimony is transmissional rather than generative» like perception (Audi 2015, p. 239).

[2] Audi 2015, p. 249.

1.2 Foundational and propositional nature of testimony

Testimony is conceived as a «social foundation of knowledge» (Audi 2015, p. 238). Testimony is *social* because it is based on a communicative interaction; *foundational* because its format is *propositional*, i.e., testimony's contents are taken to be as «affirmational saying» (Audi 2015, p. 237). Through testimony, it is impossible to «produce conceptual learning without producing some beliefs» (Audi 2015, p. 231).
The contents of testimony are gained initially without justification. Briefly, some propositions that constitute testimony are known non-inferentially, «and in that sense are *foundationally*» (Audi 2015, p. 239).

> In the natural developmental order of things, content goes from the outside in, justification from the inside out. Without the conceptualization that arises from the testimonial introduction of content, there would be no internal ground sufficiently rich to nurture justification. Particularly in children, *testimonially based knowledge arises inextricably bound up with conceptualization*. This external epistemic success by some testimony is a precondition for the internal evidence that *give a child justification for accepting other testimony* (Audi 2015, p. 233) [my italics]

In other words, having none justification is necessary at the beginning to form and ground primeval beliefs (in the form of *propositional attitudes*) as knowledge testimonially transmitted that is stored, memorised, and will serve to gain, accept, or reject other further and future contents of knowledge, thus, to fund justification. In this sense we might claim that *testimony yields knowledge first, and then justification*. I think this passage implicitly refers to Wittgenstein's *On Certainty* (1949/1969)[3], when the Austrian philosopher claims, for example at § 212: «Somewhere we must be finished with justification», or at § 253: «At the foundation of well-founded belief lies belief that is not founded». There are some beliefs[4] that we must accept as assumptions, i.e. that we learn and cannot investigate further; beliefs that «are exempt from doubt, are

[3] In this essay, Wittgenstein dedicates many reflections to child learning that would deserve to be investigated in more details. At the end of the present chapter, I will mention other passages that are compatible with my view of natural pedagogy theory.

[4] Wittgenstein uses the term "propositions" that is perfectly compatible in this context.

as it were like hinges on which [other beliefs/propositions] turn» (§ 341), and «If I want the door to turn, the hinges must stay put» (§ 343). In this sense, Wittgenstein can affirm that «certain propositions seem to underlie all questions and all thinking» (§ 476). The basic-beliefs acquired without justification represent a treasure-trove of knowledge to draw from in order to evaluate the trustworthiness of the testimony rather than *only* the honesty of the attester.

The initial success of testimony in producing knowledge without the commitment of justification serves as a precondition for concept acquisitions and future learning processes (among them language acquisition). In Audi's view, the absence of justification (i.e., absence of evaluative competencies) entails non-inferential processes. Therefore, he holds that «testimony-based knowledge is non-inferential», in the sense that «it constitutes knowledge not based on […] justifiably believed premises» (Audi 2015, p. 241).

Audi assumes that to gain testimonially based knowledge, it is normally required only «having no reason to doubt the attester's credibility» (Audi 2015, p. 232). He hypothesised that «knowledge arises before justification» (Ibidem), namely he posits an autonomy of knowledge from justification and he presupposes that «testimonially based knowledge seems to be part of the cognitive foundation from which children acquire the evidence they need to achieve justification for accepting testimony» (Audi 2015, pp. 232-233). I suggest applying for children what Audi defines «moderate inferentialism», by which «once the recipient concludes that the attester is credible, the information-receiving door is opened» (Audi 2015, p. 240).

1.3 Moderate inferentialism and local reductionism for the credibility of the attester

Imagine that the spontaneous infants' inference could be as follows: "Since the person in front of me is so reliable, all information she/he is communicating me is important and ought to be learnt". Such inference can be adequate to express the epistemic trust, namely, the cognitive bias triggered by the ostensive knowledge transmission in a pedagogical context. In natural pedagogy theory the *credibility* is given through the communicative modalities of transmission, i.e., *how* the testimony is

communicated, namely through ostensive communication. Natural pedagogy predicts that the ostensive nature of communicative transmission tend to drop any need for scrutiny and allows fast and efficient learning. However, as I will show in more details through a series of evidence, children's trust in attester is not granted, rather it is a result deeply affected by a dynamic and conflicting balance between the inclination to trust blindly induced by the teacher's communicative approach and the increasing competence of reasoning together with abilities to read others' intention and mental states.

Given such contextual variables due to relational communication, I prefer to rely on Fricker's «*local reductionism*». In face-to-face communicative interaction, Fricker (2006) proposes that only a *local reduction of testimonial justification* is necessary. Against the antireductionism - according to which when the recipient receives the relevant testimonial utterance, she must assume that the speaker/attester is *always* trustworthy (and rational) – and, at the same time, against the global sceptical reductionism[5], Fricker deems that «the subject needs not establish the trustworthiness of testimony in general, but only to establish the trustworthiness of the relevant speaker on the relevant occasion and with respect to the relevant topic» (Michaelian 2010, p. 400).

Local reductionism implies, thus, "local trust". The primary element for triggering the process of testimony is *trust*. As Adler claimed: «testimony only succeeds if there is trust» (Adler 1994, p. 264). The recipient's trust in the attester opens swiftly the door to transmitted knowledge. Following Gergely's suggestion about the gradual

[5] Reductionists view testimony as source of justification since they claim that knowledge entails justification *only* when «the recipient has an appropriate degree of independent justification for taking the attester to be credible» (Audi 2015, p. 253). As Jennifer Lackey (2006, p. 160) puts it, according to reductionists, «testimonial justification is *reducible* to sense perception, memory, and inductive inference, [while] non-reductionists maintain that testimony is *just as basic* epistemically as these other sources». If the reductionist position is historically represented by Hume - who claimed that we can rely on others because *experience* has previously shown them to be reliable sources of knowledge -, the anti-reductionism is historically represented by Reid, who notoriously wrote: «The wise and beneficent Author of nature, who intended that we should be social creatures, and that we should receive the greatest and most important part of our knowledge by the information of others, hath, for these purposes implanted in our natures two principles that tally with each other. The first of these principles is a propensity to speak truth [...] [the second] is a disposition to confide in the veracity of others, and to believe what they tell us» (Reid 1764, pp. 238-240, quoted by Origgi 2004, p. 66).

inhibition of epistemic trust in favour of a more significant refinement of mindreading competences, the increase of epistemic vigilance (Sperber et al. 2010) may be due to a more sophisticated monitoring process focused on both the honesty of the attester, and her alleged competences. From a certain developmental phase, we may claim individuals form testimonial beliefs using a monitoring process, that involves (unconscious) cognitive activity to determine sincerity and competence, as suggested by Fricker (1994, p. 150). However, Fricker conceives honesty and competence as default settings within the monitoring process, whereby «the subject assumes trustworthiness only given that she has actively attempted to determine whether the speaker is trustworthy» (Michaelian 2010, p. 406). According to Fricker's (1994) conception, trustworthiness encompasses the notion of sincerity (or honesty) and competence of the attester. Thus, we will have that if the subject believes i) that the agent affirms P, ii) that the agent is then competent with respect to P, and iii) that she/he was then honest with respect to P, the monitoring process outputs a belief that P (Ibidem).

1.4 The circularity of trustworthiness

All of us «must evaluate the trustworthiness of others to evaluate their testimony» (Lehrer 2006, p. 145). However, the search of the reasons why one must hold for trusting (or distrusting) testimony leads to a circular loop of trustworthiness. In fact, «if I must evaluate the trustworthiness of others before accepting their testimony, [then] I must evaluate my own trustworthiness by appeal to the testimony of others [. Therefore,] I am trustworthy in evaluating others, […] only if I am trustworthy in evaluating the testimony of others concerning the correctness of my evaluations» (Ibidem). Leher suggests that the solution to the circularity problem depends on recognising that «I must accept that I am trustworthy in my evaluations of people, myself included» (Lehrer 2006, p. 154).

I suggest that such self-inclusion, that must be automatic, can be ensured by the normative dimension of social learning *based on testimony*. The acceptance is given by the normative interpretation of asserted and transmitted message. In natural pedagogy, trustworthiness of testimony is combined with reliability of knowledge source. An implicit acceptance agreement is ratified between counterparts, permitting in this way

not only the mere communicative transmission but also the transformation of conveyed information into (propositional) knowledge. This informational baggage, in turn, constitutes those *common ground* beliefs that serves both to evaluate (i.e. provide justification for) further testimonies, and to support those pragmatic presuppositions useful for sustaining social communicative interchanges. The former point entails that justification of testimony is not foundational, because the source of the justification is «the background system to which we have referred and which we use to meet objections to the testimony we accept» (Lehrer 2006, p. 151). Therefore, at the beginning of our belief foundational process we have *acceptance* of testimony based on an act of *trust*. Without at least some measure of trust, «one might not acquire testimony-based belief at all» (Audi 2015, p. 256). One should trust, first. This seems to be the epistemological condition for transmission of knowledge, and from psychological (and epistemological) point of view in line with natural pedagogy, *trust is never granted absolutely*.

2. Learning by trusting

Trust is not a mere output of a gullibility attitude, rather it is the result of an interpersonal relation, and at least of a communicative relation. Epistemic authority and trust are deeply connected (Origgi 2004). One of the challenge for the ongoing cognitive developmental studies is to determine the maximum and the minimum threshold of trust. Having trust involves conscious and unconscious commitment of an evaluative first person. I am going to show that young children are not subjects totally devoid of evaluation capacities.

Evaluative first person emerges overwhelmingly little by little. Some critics of natural pedagogy, as Nakao and Andrews (2014) for example, accuse natural pedagogy of depicting pedagogical stance as a passive attitude that involves gullible infants who trust anyone. In this way, I hold Nakao and Andrews exaggerate some descriptive aspects of the theory undermining the *relational component* that the theory defines as a premise. The risk is to flatten the theory on the apparent congruence with Thomas Reid's claim about human *a priori* trust in testimony (see footnote n. 4 about the reductionism in the present chapter). According to Reid (1764/1997, p. 197), credulity

is intended as a «gift of nature» and is especially strongest in childhood. As Reid wrote (1983, p. 281): «The wise Author of nature hath *implanted in the human mind a propensity to rely upon human testimony before we can give a reason for doing so*. This, indeed, puts our judgment almost entirely in the hands of those who are about us in the first period of life» [my italics]. Koenig and Harris (2007, p. 270) noticed that «on Reid's view, testimony is treated like first-hand experience». On the contrary, following Elizabeth Fricker (2006), testimony allows the acquisition of knowledge through trust, and such «*knowledge* […] is always and necessarily *knowledge at second-hand*» [my italics] (Fricker 2006, p. 593). Fricker's points out also the normative dimension of «knowledge gained through trust in testimony» that is consistent – I suggest - with the predictions of natural pedagogy theory: «when a hearer forms belief in what she is told through trust in the teller, […] *she takes the teller to be expressing knowledge*, and this normative commitment is an *essential part* of the hearer's grounds for her belief» [my italics] (Fricker 2006, p. 593).

I suggest that such notion of testimony (as «knowledge gained through trust») is entailed by the theory of natural pedagogy. Moreover, following Sperber and collegaues (2010) I hold that, on the one hand, the development of metarepresentation skills fosters epistemic vigilance enabling the child to evaluate the *reliability* of the source. On the other hand, early inferential capacities make infants able to analyse (to some extents) the *trustworthiness* of the message. Therefore, young children are not always the naïve credulous subjects thought so far. In fact, as Harris (2007) showed three- and four-year-olds are able «to monitor the accuracy or knowledge of informants, including those that are familiar. They prefer to seek and endorse information provided by someone who has proved accurate in the past rather than someone who has made mistakes or acknowledged ignorance» (Harris 2007, p. 135). Furthermore, infants seem to be endowed with logical inferential skills (Cesana-Arlotti et al. 2018) and can interpret actions according to the principle of rationality (Csibra et al. 2003).

The biases of natural pedagogy are triggered in virtue of certain situations that allow rapid and efficient knowledge transmission. Such conditions are characterised by a friendly approach through amodal communicative channels based on the use of special ostensive signals. The communicative relationship established in this way falls under the notion of testimony which in turn is hinged on the concept of *trust*. Over time,

children learn to ration out this trust thanks to the improvement of metarepresentation capacities. Thereby, the communicative relationship that grounds learning process goes on along a double evaluation track: the reliability of the information source, and the trustworthiness of the message.

2.1 Trustworthiness and reliability of the teacher (attester). The conquest of trust

Reliability permits a confident predisposition to imitate the attester's performance. In their findings, Zmyj and colleagues noticed that «infants' imitation of the novel action was influenced by the model's previous reliability; they copied the action more often when the model had been reliable» (Zmyj et al. 2010, p. 208). The reliability of the source, conquered in virtue of the epistemic trust, represents likely an assurance for the benefit the information transmitted (say *p*), namely, an assurance of the validity and the importance of *p* from the recipient's perspective. However, I deem this issue deserves a clarification that regards the relationship between the *reliability of the communicative source* (i.e., the knowledge holder) in terms of *trust*, and the *trustworthiness* of the knowledge transmitted (i.e., in the attestation) (Audi 2015, p. 255; Lehrer 2006). Have infants cognitively to support a sort of what Audi (2015, p. 250) calls «presupposition of trustworthiness»? The trustworthiness is a matter of the credibility that regards directly the testimony.

> Trust is fulfilled when its object meets certain expectations in those who trust; trustworthiness is fulfilled when the subject, the trustworthy person, meets (or would meet) certain expectations (or hypothetical expectations) on the part of someone who trusts the person (Audi 2015, p. 255).

Depending on the age, the young evaluator will be able to monitor the informant judging her/his trustworthiness, the accuracy of the performance, her/his social prestige, and the conformity to the acquired shared social norms. The first target for a social learner is to *select reliable informant*.

2.1.1 Verifying reliable sources

Findings attested that children manage to select their source of information and prefer to learn from individuals who show to be reliable (Harris 2007; Mascaro&Sperber 2009; Palmquist et al. 2016; Poulin-Dubois et al. 2011). Preschoolers usually prefer trusting adult over peer informants, even if Jaswal and Neely (2006) have illustrated in their study that children may prefer a peer informant if she shows to be more accurate.

The selective trust is made also on the base of an informant's social conformity, as claimed by Harris and Corriveau (2011), according to which children tend to «trust informants who are culturally prototypical» (Harris&Corriveau 2011, p. 1183). Therefore, the young apprentices seem to be circumspect as if they knew that «not all information conveyed by others is accurate or worth learning» (Poulin-Dubois& Brosseau-Liard 2016, p. 60).

Poulin-Dubois, Brooker and Alexandra Polonia (2011) assumed that the use of positive emotional referential signals create a special faithful relation between infant and the attester of information[6]. The researchers investigated whether «the past reliability of a person's emotional signals influences infants' willingness to imitate that person» (Poulin-Dubois et al. 2011, p. 303). They utilised an experimental set-up very akin to Gergely et al. (2002) in order to test imitation task of odd actions like turning a lamp on using the forehead, and noticed that «infants with exposure to a reliable model were more inclined to imitate the model's novel actions» (Poulin-Dubois et al. 2011, p. 306). Harris and Lane (2014) summarise the experiment as follows:

> «Infants ranging from 13 to 16 months[7] watched as a person looked inside a box and produced a positive affective response, ''Wow!'' In the reliable-looker

[6] See in Chapter 3 Egyed et al.'s (2007; 2013) experiments (based on Repacholi (1998) paradigm) that recall the same communicative function. According to the "natural pedagogy" stance on learning, children will assume that adults act with rational and efficient means to demonstrate new and culturally relevant information (Gergely&Csibra 2005; Gergely et al. 2007) through emotional expressions. See also Landrum et al. (2016).

[7] The mean age for 60 infants who participated the study was 14.39 months (Poulin-Dubois et al. 2011, p. 305).

> condition the box contained a small toy but in the unreliable-looker condition the box was empty. Infants then saw that person perform a novel action: when seated at a table with a lamp, the person leaned over, pressed the lamp with her forehead, and turned it on. Meanwhile, her hands rested on the table, so it was clear that she was deliberately not using her hands[8]. The lamp was then placed in front of the infant, and the experimenter said, ''Now, it's your turn.'' Infants in the reliable-looker condition more often turned the lamp on with their foreheads—whereas infants in the unreliable-looker condition more often turned on the lamp with their hands, as they would normally» (Harris&Lane 2014, p. 448).

The results show that infants' previous experience with a reliable source of information influences their consequently imitative behaviour of the model's demonstration. «Only 34% of the infants in the unreliable group imitated the model using their forehead, in contrast to 61% of the infants in the reliable group […]. Of all the infants who used their forehead at least once […] 65% belonged to the reliable group» (Poulin-Dubois et al 2011, p. 307).

The year before, Zmyj and colleagues (2010, p. 208) had arrived at the same conclusions investigating whether 14-month-olds were able to «(a) imitate instrumental actions and (b) adopt the individual preferences of a model differently depending on the model's previous reliability». Such reliability was grounded on a judgment of competence or incompetence that the model showed acting on familiar objects. Zmyj and co-workers too used the unusual action of lighting a box with the forehead performed by a demonstrator. He was presented in a series of videos in which he did wear a pair of shoes either correctly (reliable condition) or incorrectly (unreliable condition). «As soon as the test videos ended, the experimenter entered the room, placed the apparatus [box shaped lamp] used in the video, told infant "Now you can play with it", and left the room» (Zmyj et al. 2010, p. 211). In the reliable condition, 59% of

[8] The procedure changed in respect with Gergely et al.'s (2002) task, specifically the hands-free condition, in order «to see whether infants' prior knowledge of an experimenter's credibility in the emotional referencing task would influence their willingness to imitate her behavior in a novel action task. Therefore, this task was always administered after the emotional referencing task. […] As [the demonstrator] performed [the] head-touch [action], her hands were always placed flat on the table, on either side of the light. This action was repeated for a total of three times» (Poulin-Dubois et al. 2011, p. 305).

infants imitated the novel action, while in the unreliable condition only 30% of infants. These findings «demonstrate that 14 month-olds can take into account a model's previous reliability when socially learning from him[/her]» (Ibid., p. 218).

2.1.2 Modalities of communication influence infants' evaluation of informants

The communicative approach is determinant also with children and not only with infants. Landrum and co-workers (2016) found that 4- and 5-year olds evaluated (i.e. made epistemic inferences about) their informative source in respect both of the niceness and the accuracy of the attester. When 96 preschoolers were asked to attribute knowledge and behaviours to a nice, a mean and a neutral informant with different degrees of expertise, children attributed more knowledge to nicer informants, but they also attributed more knowledge to expert informant. This study indicates that «children have a complex view of the relationships between observable informant features and informant mental states, at least when evaluating informant knowledge» (Landrum et al. 2016, p. 711). Equally, Palmquist and Jaswal (2015) claimed that: «the scope of inferences children draw about the knowledge of informants can be influenced by the modality in which they communicate» (Palmquist&Jaswal 2015, p. 178). These findings seem to confirm the parallel development of the two tracks of evaluation: one is underpinned by mindereading system, the other one is implemented by inferential strategies that compare and compute previous beliefs. In this regard, Michaelian (2012a; 2012b) proposes a «metacognitive epistemology framework (MEF)», according to which the metacognitive monitoring of one's own memory (and beliefs stored) plays a more significant role than monitoring of others' mind in «ensuring the reliability of testimony» (2012b, p. 481). In other words, the memory store represents the information source, while metamemory[9] processes monitor, retrieval and determine the endorsement or the rejection of the novel belief» (2012b, p. 501).

[9] Metamemory is a type of metacognition that refers «to knowledge about one's memory capabilities and strategies that can aid memory, as well as the processes involved in memory self-monitoring» (Pannu&Kaszniak 2005, p. 105). Probably, in this case underlined by Michaelian, metamemory is not necessarily requested. I thank Cristina Meini for suggesting me this caution.

2.1.3 Strengthening of monitoring capacities

The attested «age-related improvements in children's ability to trust the more reliable informant are likely due to changes in the ability to differentiate, encode and monitor multiple sources of information» (Koenig&Harris 2007, p. 274; Koenig et al. 2004). In the study of Koenig and colleagues (2004), 3-4 year-old children «were given an opportunity to assess the comparative reliability» of two unfamiliar adults who named a series of four familiar objects (Harris&Corriveau 2011, p. 1181). One informant termed always all objects correctly, so when a ball was presented, she said: "That's a ball". The other adult, instead, «named all the four objects incorrectly. Presented with a ball […], she said: "That's a cup"» (Ibidem). Since all the children knew the names of the objects, they could judge the accuracy of informative source and understand that one informant was always right, and the other one was always wrong. In this way, when «unfamiliar objects were presented - whose names were not known to the children - they preferred to ask for information from the accurate as opposed to the inaccurate informant»[10] (Ibidem). A subsequent research, conducted by Pasquini and coworkers (2007), showed analogue responses about selective children's trust in accurate informants although the two adult informants provided sometimes correct and sometimes incorrect claims. In particular, «children watched one informant who was predominantly correct (75% of the trials) and another who was predominantly incorrect (75% of the trials) during induction. Even though both informants had been sometimes right and sometimes wrong, children still went on to invest greater trust in the more accurate of the two» (Harris&Corriveau 2011, p. 1181). Moreover, it is noteworthy that although children typically prefer trusting an adult over a peer informant, Jaswal and Neely (2006) noticed that preschoolers may prefer peer informant, if she shows to be more accurate.

[10] See experimental results reported in Koenig and Harris (2007, pp. 273-274) and in Koenig, Clément and Harris study (2004) where 3 and 4 year-olds «were presented with two informants who were consistently accurate and inaccurate regarding familiar object names», and after three trials «children were asked for an explicit judgment "Which one of these people was not very good at naming things?". Then at the end of the study, after receiving no new information about accuracy, children were asked this question again as a check of their memory. […] 50% of the 3-year-olds and 70% of the 4-year-olds performed perfectly on the explicit judgment probe» (2007, pp. 273-274).

«Children select whom to approach for information and whom to believe […] depending on [children's] history of interaction with those informants» (Harris&Corriveau 2011, p. 1180). Therefore, if children cannot initially discriminate *what* to believe, «they are nonetheless quite selective in choosing *whom* to believe» (Ibidem). We must abandon the «conception of young children as prone to indiscriminate trust», in favour of a conception that sees children enable to «trust some informants more than others» (Harris&Corriveau 2011, p. 1179). On the one hand, in Harris and Corriveau's investigation, children were able to «keep track of the history of potential informants» endorsing those claims made by someone who had provided «reliable care or reliable information in the past»; on the other hand, children monitored «the cultural standing of potential informants» endorsing those claims made by someone «whose behaviour abide[d] by, rather than deviating from, the norms of their group» (Ibidem).

The infants' natural inclination to trust does not mean a priori credulity. Credulity has a cost because it leads young children to be deceived in virtue of those communicative biases that allow them to acquire swiftly knowledge, as Jaswal and colleagues (2010) showed. In one of their study, for instance, «children who could both see and hear a deceptive speaker were more likely to be misled than those who could only hear her» (Jaswal et al. 2010, p. 1541). This can be read as a further demonstration of how the epistemic trust bias triggered by pedagogical conditions is robust. It is plausible to think that children look for and expect true information rather than false information. Monitoring the accuracy of informants should increase the probability that children learn from *reliable*, *honest* and *truthful* informants.

> «[Children] endorse claims made by informants who respect rather than deviate from the morphological rules of their language. They endorse demonstrations of tool use by models who speak with a native as opposed to a foreign accent. They endorse claims made by informants who elicit bystander approval rather than disapproval. Finally, they endorse claims made by members of a consensus rather than those made by lone dissenters» (Harris&Corriveau 2011, p. 1183).

Further findings indicate that children (from 3- to 8-year-old) «may be inclined to doubt speakers who make claims they cannot verify themselves, as well as a

developmentally *increasing* appreciation for speakers who make general claims» (Koenig et al. 2015, p. 22). Indeed, preschoolers seem to be aware of the risk of misinformation demonstrating selective learning (e.g., Koenig et al. 2004). At the same time, it is true that in many contexts children show striking credulity (e.g., Jaswal et al. 2010). Stephens and Koenig (2015, p. 182) attempt to «explain these divergent patterns by examining the possibility that errors for semantic information, a type of information that is typically generalizable and difficult to verify independently, promote greater vigilance than errors for episodic information, which is often event-specific and independently verifiable». In their experiment, 3- and 4-year-olds did see two speakers perform «correctly or incorrectly about object labels (Semantic condition) or locations (Episodic condition)». The test revealed that children exposed to semantic inaccuracy were more vigilant than children initially exposed to episodic inaccuracy. According to these results, the authors «speak against a homogeneous treatment of testimony and suggest that preschoolers' testimonial vigilance varies according to the content of speakers' errors». As Sterelny recognises too, infants «assess the reliability of sources, as well as the plausibility of messages» (Sterelny 2012, Chapt. 6.1).

In addition, in this regard, Yi and Low (2018) investigated the power of selective trust based on the informant's past history of accuracy, by presenting unexpected (but not completely unbelievable) testimony that contradicted 3- and 4-year old children's initial beliefs. The researchers found that «children expressed a greater tendency to override their initial judgments and endorse the unexpected testimony from a previously accurate informant than from someone who had consistently made naming errors» (Yi&Low 2018, p. 1).

2.2 *Evaluating trustworthiness*

In the present paragraph, I am going to report a recent study that shows how logical capacities allow to judge events and create expectations. I hold that the following experiments are crucial for my argumentation because they endorse the hypothesis that infants are not so credulous in respect of the content of the message. This suggests that natural pedagogy's biases work dynamically cooperating with such inferential processes.

Cesana-Arlotti and colleagues (2018) wondered whether 12-19-month-olds are able to use disjunctive syllogism, also known as process of elimination (or *modus tollendo ponens*) that follows the argument of the form: A or B, not A, therefore B.

> [Cesana-Arlotti's team] measured infants' looking at computerized vignettes in which two different objects (A and B) were shown being hidden behind a wall. Infants watched as a cup scooped one of the objects from behind the wall, and then came to rest next to the wall—critically, only the topmost edge of the contained object could be seen peeking out of the cup, such that infants could not tell for sure whether the object was A or B. At this moment, infants could have formed a disjunctive thought—for example, "either the object in the cup is object A or it is object B." Next, this ambiguity was resolved: The wall dropped to reveal that object A was behind the wall, but the contents of the cup remained hidden (Halberda 2018, p. 1215).

This was the crucial moment for verifying whether toddlers made actually the inferential process of elimination, called by Cesana-Arlotti et al. (2018, p. 1263) «potential deduction phase». Indeed, infants had the opportunity to draw the following inference "because object A is not in the cup, object B must be in the cup." (Halberda 2018). The researchers identified «markers of inferential activity by examining the dynamics of oculomotor responses during inference making. Last, [they explored] stability across development by comparing the oculomotor responses of infants, toddlers, and adults passively looking at nonverbal scenes that potentially involve logical inferences» (Cesana-Arlotti et al. 2018, p. 1263). They found that infants looked longer when the unexpected object A emerged from the cup. Such outcome means that their expectation based on logical inference was violated. However, this striking result was achieved by VOE method that measures a response «after a conclusion has been reached» (Cesana-Arlotti et al. 2018, p. 1264), so it does not characterise «the unfolding of an inference in the infant mind». For this reason, researchers measured «oculomotor responses during the potential deduction phase» (Ibidem). This time, in a novel experiment akin to the previous one, «infants already know which object is in the cup before the potential deduction phase», and also in this case they «looked longer at an outcome inconsistent with the identity of the (known) object in the cup», and the

experimenters noticed that «during the potential deduction phase, the infants' pupils dilated more when the scene licensed an inference than when it did not, suggesting increased cognitive activity possibly due to inference-making» (Ibid., pp. 1264-1265)[11].

> Only when the potential deduction phase afforded an inference did higher pupil dilation and visible object-to-cup shifts contribute to predicting success at identifying inconsistencies in the later outcome phase (Ibid., p. 1265).

The employ of oculomotor markers indicates more clearly that preverbal infants are able to deploy efficiently «logical procedures to process the components of an unfolding scene» (Ibidem). As Cesana-Arlotti and colleagues emphasise:

> This empirical evidence is directly relevant to the old, yet still fundamental questions debated between [Chomsky,] Fodor and Piaget[12]. Logical representations that are crucial components of infants' natural hypothesis-testing attitude are available when *infants start projecting and testing hypotheses about the world*. Such representations may consist of nonlinguistic but *fully language-like structures*, or they may piggyback on sophisticated object representations that can track object identities in ambiguous situations [my italics] (Cesana-Arlotti et al. 2018, pp. 1265-1266).

These final considerations are congruent with what I exposed in the prior chapter about the representational structure of the information processed by infants, and it is worth noting that, like in the case of spontaneous false belief attribution that children explicitly master later, also for verbally expression of disjunctive reasoning occurs such developmental dissociation.

The logical capacities shown by this finding endorse the hypothesis that children begin preciously to evaluate the events of the world and to refine those skills that will be soon useful for judging the trustworthiness of testimony, or when children will have

[11] Also adults dilated pupils during potential deduction phase like infants, but they differed only in the speed of the markers (Cesana-Arlotti et al. 2018, p. 1265).

[12] The authors refer to Piattelli-Palamarini (1980) in which Piaget argued that logic in human infant mind was the pick of a long development path extending into the adolescence.

to tackle those events that might contradict their background knowledge. As in the two experiments created by Van de Vondervoort and Friedman (2017, p. 15), in which «3- and 4-year-old children spontaneously provided corrections and protested pretense scenarios in which animals produced sounds typical of a different species». The researchers found that «children even provided protests and corrections when the experimenter signaled that the pretense might include unrealistic elements». This study confirms that «children use their general knowledge to generate and interpret pretense, and show that they attach *normative force* to the content [even also] of pretend play» (Van de Vondervoort&Friedman 2017, p. 15). Therefore, I suggest that, on the one hand, children may acquire new knowledge normatively through natural pedagogy model, intended as a form of testimony transfer, and on the other hand, children apply inferentially such *stable* knowledge to judge *novel* sources of information.

3. What kind of pupils for natural pedagogy theory?

According to Harris and Lane (2014), Heyes (2016), Nakao and Andrews (2014) natural pedagogy model depicts infants more as attentive and receptive pupils rather than «self-directed» pupils who actively seek «to remedy it by consulting nearby adults for relevant information» (Harris&Lane 2014, p. 454). However, I hold that such critique is not correct, because natural pedagogy focuses on knowledge transfer and how such transfer can occur. The theory describes a communicative relationship which is, by definition, an exchange influenced and determined by the principle of epistemic primacy that portrays infants - as someone has written - «avid seekers of information provided by others» (Poulin-Dubois et al. 2010, p. 303). How do children decide which actions to imitate and learn? Imitation occurs in different social learning contexts «dependent upon the degree of object affordance, task novelty, and task complexity» (Dickerson et al. 2013, p. 719). There is «a flexibility in children's copying behavior», whereby sometimes children copy exactly, «and sometimes copy selectively» (Over&Carpenter 2012, p. 182; Over&Carpenter 2013).

According to Buchsbaum, Gopnik, Griffiths and Shafto (2011, p. 331), «children combine rationally multiple sources of information», and, if on the one hand they can «learn from contingencies between action sequences and outcomes across repeated

demonstrations», on the other hand, «children interpret the same statistical evidence differently when it comes from a knowledgeable teacher versus a naïve demonstrator» (Buchsbaum et al. 2011, p. 331). In the pedagogical condition, children tend to 'overimitate', i.e., to reproduce the entire sequence of actions more faithfully. This suggests that «children attend to both statistical and pedagogical evidence in deciding which actions to imitate» (Ibidem). The pedagogical communication represents «a form of epistemic cooperation that can transfer relevant episodic information about specific referents as well as generic knowledge about referent kinds» (Gergely 2011, p. 81). In other words, there is a need of knowledge that triggers the research of information expressed by the utilisation and recycling of communicative means.

To sum up, social learning encompasses (and is accompanied by) the development of two innate inferential tracks of evaluation that must select what and from whom to learn. «Learning from others requires integrating reasoning about an informants psychological properties, such as knowledge and intent, with reasoning about the implications of the data the informant chooses to present». Such two reasoning paths are interrelated, and «guide learners when acquiring knowledge about the world» (Landrum et al. 2015, p. 109).

4. *Low-level and high-level inferential process for triggering and sustaining natural pedagogy system*

I propose that the inexorable development of such routes will conduce to inhibit natural pedagogy social learning, because they are conflicting with triggering conditions and the operative biases. The increase of epistemic vigilance inhibits the fast and low-level cognitive processes that allow, in virtue of the eliciting presence of ostensive cues, to recognise communicative intentions and trigger referential expectations. At the same time, the predominant activity of mindreading system undermines progressively not only the epistemic trust, but also the high-level processes that implement natural pedagogy, namely the assumption of universality. Then, the more sophisticated evaluative competences restrain the *genericity* interpretation bias too. However, I would like to point out that such ongoing restriction of natural pedagogy's application throughout childhood is carried out along a *developmental trajectory*. Thus, I deem too

strong the claim by Nakao and Andrews (2014, p. 465) according to which «children engage in *active learning* by *deciding which models to imitate* in order to gain knowledge about their physical world and the norms of their culture» [my italics]. Here, I do not want to deny the active learning aspect abovementioned, rather what I have tried to explain is that although infants are not passive and totally credulous pupils, at the same time they are not always ready to examine and select the benevolence and the expertise of the demonstrator in front of them in force of child-directed teaching employed by natural pedagogy system.

Infants and children, as information seekers, take advantage of learning opportunities favoured and conveyed through ostensive signals. Such communicative cues represent the easy access to the pedagogical condition. I wondered (and discussed in the second chapter above all) whether the ostensive signals provoke communicative intentions recognition and induce referential expectations in force of mindreading competences at work. My position rests skeptical in this regard on the basis of available findings. I propose the revision of natural pedagogy for what concerns the higher cognitive process that universality assumption requests. In my opinion, it is not enough to claim – as Csibra and Gergely did - that the addressee of pedagogical knowledge transmission «has the *default expectation* that the content of the demonstration represents shared cultural knowledge» (Csibra&Gergely 2011, p. 1150). By using the definition «default expectation», we take the risk to underestimate the cognitive cost of the presumption of omniscience, that has to be interpreted as knowledge (or belief) attribution. Such attribution competence is the result of a conflicting and dynamic cooperation between natural pedagogy and mindreading system. Around the fourth year of age such conflict is more evident, but it does not imply the end of the natural pedagogy system functioning, rather a faster restriction of its application.

5. The theory of natural pedagogy under attack

The most part of attacks (e.g. Heyes 2016; Moore in press[13]; Nakao&Andrews 2014) that natural pedagogy model received regards the determinant role played by «the ostensively communicated manifestations» (Topál et al. 2008, p. 1832). As I hope I

[13] I thank Richard Moore for allowing me to use his draft paper.

have sufficiently highlighted in the second chapter, infants (from birth) prefer communicative faces, and in general communicative approaches, even if these come from unfamiliar persons. During the ontogenetic development, the amodal strength of ostensive communicative channels changes dramatically. If infants prefer eye contact, smiling and simple vocalisations, children are more confident with articulated IDS. However, it is worth noting that very early infants appreciate more native speakers than foreign speakers. In this regard, Kinzler, Dupoux, and Spelke (2007) showed «that 6-month-old infants spend more time looking at someone who has previously spoken their native language than someone who has previously spoken a foreign language»; and «10-month-olds prefer to take the object offered by [a] native language speaker» rather than by a foreign language speaker (Buttelmann et al. 2013, p. 423).

Csibra (2010) lists specific ostensive signals but there could be others. Adults use motherese as an affective form of approaching babies also when they do not want to communicate or teach anything. There is no reason to think that such communicative practise is evolved functionally for serving *only* a teaching role. It seems to me that Nakao and Andrews, Richard Moore (2016a; in press) and Heyes (2016), read the role of ostensive signals within natural pedagogy framework as an absolute dictatorship whereby the mere presence of the signals must always indicate a teaching context that infants must decode. It seems to be valid the inverse too, namely that teaching context has to be accomplished almost exclusively in virtue of ostensive signalling (Moore 2016a). This interpretation risks to be reductive and mislead the role of ostensive communication within natural pedagogy system. I think that Csibra and Gergely (2006) emphasise the adaptive functional role of ostensive signals in an evolutionary perspective in order to explain the ease of learning and conservation of pragmatic knowledge shown by infants. Imitation by observation is not enough to take into account the strength of the intergenerational knowledge transmission. The direct involvement of infants with the action they must interpret (and that involves an external social and relevant object of the world) can represent the best modality of learning. Under this light, the functional role of specific communicative means can be better understood. Ostensive signals are gestures and words that can be used for several purposes. One of these is to manifest the intention to do something, to say something. If we focus on the cognitive side of the matter, which is the core of our investigation, the

functional role of ostensive communication is to convey communicative intention that when arrives at destination triggers in the young addressee a referential expectation. This the core of the pedagogical stance. Once established such connection, and the referential expectation is triggered, the demonstration can begin, and the adult agent can transmit the content of the message. I would be tempted to claim that without communicative intention we have no teaching act, but I would be wrong. In fact, Kim Sterelny (2012) proposed a cultural learning model that does not need an explicit form of intentional teaching. This helps us for clarifying the contextual social limits of natural pedagogy account.

Sterelny's cultural learning model converges with the Csibra-Gergely model, as Sterelny acknowledges even if I suppose he misleads natural pedagogy theory when he claims that it «focuses *wholly* on the role of language» (Sterelny 2012, Chapt. 6.4). Sterelny (2012, Chapt. 2.3) provided the «apprentice learning model», that is a hybrid learning model: it combines «information from the social world with information from the physical-biological environment. It is learning by doing». According to Sterelny (2012, Chapt. 2.3), the reliable transmission of know-how begins with the observation of adult activity, without direct adult teaching («explicit instructions» and «formalised institutions») or special adaptations for social learning in the children, who engage exploratory trial-and-error learning in their environment that is organised by adult experts combining «task decomposition» and «ordering skill acquisition». In this way, adults can structure the learning environment of the children «without much explicit teaching». In brief, if the role of explicit teaching - like in the case of pedagogical relations - is quite limited, adults can «structure and engineer the learning environment, even without explicit teaching» (Sterelny 2012, Chapt. 2.3)[14].

[14] Shneidman and Woodward raise the question «whether child-directed interactions are the critical route for supporting cross- generational transmission of cultural information», and comparing data «from U.S. and European children, as well as from communities where directed interactions with young children are rare», the researchers doubt «that directed interactions provide automatic and innate informational value for learners» (Shneidman&Woodward 2015, pp. 1, 3).

5.1 Explicit teaching through ostensive cues: further remarks

Ostensive signals are like a beautiful car that carries friendly communicative intentions. In order to arrive at destination, in fact, ostensive signals must be nice and positive. Landrum and colleagues (2016) speak about «niceness». In this sense, we can claim that ostensive signals are *means to gain epistemic trust*. As I have argued so far, without trust, testimony is not possible. The *niceness* and the positive expressions of ostensive signals convey efficiently the communicative intention predisposing infants to receive information that ought to be taught.

Under the lens of the epistemology of testimony, the ostensive communication may guarantee, or at least elicit, the *reliability* of the informative source. Without such reliability the trigger of pedagogical stance cannot occur. This means that even in presence of ostensive signals, the pedagogical relation cannot be established if the reliability of informative source is undermined. The abovementioned results of Poulin-Dubois et al.'s (2011) and Zmyj et al.'s (2010) studies have to be interpreted from this perspective. I reject the interpretation provided by Nakao and Adrews (2014), who interpret them as evidence that natural pedagogy depends only on the mere presence of ostensive communication. On the contrary, we can confirm that natural pedagogy is implemented by ostensive communication that occurs in a local context and has the function to convey communicative intentions positively for attesting the reliability of the source.

Clément and Dukes recognise that: «we learn a lot by being involved in deep *affective* communicative relationships […] But such ostensive communications *are clearly not* the only way to learn about our social environment …» (Clément&Dukes 2017, p. 257). Furthermore, during the development other forms and occasions of social learning enrich the modalities of knowledge acquisition. However, among pre-schoolers, natural pedagogy model proves once again to be *one* of the best learning strategies for acquiring generic knowledge. Butler and Markman (2014) show how children use information about «a particular function as critical to category membership (over and above more salient features) when it was demonstrated ostensively», in order «to guide their inferences about what defines a novel category» (Butler et al. 2015, p. 479).

In the case of social norms acquisition too, pedagogical demonstrations convey generic, kind relevant information better than other forms of learning. In fact, «although pedagogy may *not be necessary* for inferences to the generic, it may nevertheless be sufficient to produce inductive inferences on which the child relies more strongly» (Butler et al. 2015, p. 476). Butler and coworkers (2015, p. 485) point out that: «Children made significantly stronger normative inferences about novel actions when they saw those actions pedagogically demonstrated, relative to seeing the identical actions carried out in a deliberate and intentional but non-pedagogical manner». Their finding suggests that normative inferences depend on the manner of the acquisition process[15]. In addition, pedagogical manner seems to be one of the best. However, for what concerns the normativity dimension tied to generalisation bias, Marion Vorms (2012) raised a disappointment about the explicative strategy provided by advocates of natural pedagogy theory relatively to the phenomenon called "*A-not-B error*".

5.2 Marion Vorms' critique

In 2008 the Hungarian team composed by Topál, Gergely, Miklosi, Erdohegyi and Csibra applied the explanatory natural pedagogy framework (and in particular the bias of *generalisation*) to the known phenomenon investigated by Piaget (1954) called "*A-not-B error*". They suggested that A-not-B error is due to «infants' pragmatic misinterpretation of the information conveyed by the experimenter», but according to Marion Vorms, Topál and colleagues failed «to specify exactly what this information consists of» (Vorms 2012, p. 526). A-not-B errors (consisting in a hide-and-search task) are mistakes performed by infants around the first year of life. The standard experiment is divided in two phases. During the habituation phase (or A-trials), «the demonstrator

[15] «Consistent with prior work (Schmidt et al., 2011), 3-year-old children showed a general tendency to jump to a normative interpretation from simply observing an intentional action performed by an unknown knowledgeable adult, protesting when a third party used the same objects to perform a similar, but markedly different action. But when they saw additional individuals continue to perform this markedly different action, children persisted in their normative protest significantly more when the original action had been explicitly demonstrated for them, compared to seeing the identical action performed in a non-pedagogical manner» (Butler et al. 2015, p. 485).

repeatedly hides an object under one (A) of two containers (A and B) in full view of the infant, who is allowed to retrieve the object after each hiding event» (Vorms 2012, p. 534). In the test phase (or B-trial) «the demonstrator places the object under container B and allows the infant to search for it. Despite just having seen the object being hidden at the new B location, infants between 8 and 12 months of age frequently look for it under container A where it had been previously hidden» (Topál et al. 2008, p. 1831).

Piaget explained the phenomenon as a supposed incomplete understanding of object permanence representation. This interpretation has been completely rejected by the numerous researches about infant's early capacity to entertain object permanence representation (see in this regard for a review Baillargeon 2008). Subsequent developmental researches focused their explanatory strategy (for a review see Topál et al. 2008, p. 1831) on action interpretation rather than object representation. It is the case, for example, of Longo and Berenthal study (2006) on the activation of the mirror neuron system. They suggested that observing repeatedly the demonstrator's hiding events at location A is sufficient for infants to code others' actions in terms of motor simulation. Instead, Topál and colleagues explained the infant's perseverative error by applying the «*the pedagogical learning stance*» since «the A-not-B task normally involves face-to-face interaction, in which object hiding is accompanied by the demonstrator's ostensive and referential signals […] such as eye contact, infant directed speech, addressing the baby by name, and pointing at and/or looking back and forth between the hiding location and the infant» (Topál et al. 2008, p. 1832). In other terms, the presence of ostensive cues induce infants to interpret the situation as a teaching context rather than a game. Therefore, within the natural pedagogy framework, the error is due to «a *pragmatic misinterpretation* of the nature of the conveyed information» (Vorms 2012, p. 534), determined by the generalizability assumption. As Topál and colleagues claim:

> «the interpretive bias of generalizability may result in a pragmatic misinterpretation of the object-hiding actions as potential teaching demonstrations. As a result, the infant would tend to infer and learn some generalizable information, such as "this kind of object is to be found in container A" or "we keep toys in container A."» (Topál et al. 2008, p. 1832).

In this case, it should be evident the normative strength of the information pedagogically acquired, although it does not correspond with the true message transmitted. On the basis of such theoretical assumption, Topál et al. (2008) predicted that, in absence of ostensive cues, the bias should not exert its influence. Without the cognitive 'distortion' induced by the pedagogical situation, infant should be more free to interpret the action and the game itself. However, the results of Topál et al.'s experiment can clarify only to some extent the phenomenon of A-not-B errors. Indeed, the Hungarian researches found the direct confirm that in clear pedagogical condition – i.e. Ostensive-Communicative (OC) context where a female demonstrator established eye contact with each single infant, and addressed her/him in infant-direct speech – the 86% of 10-month-olds committed errors. The percentage decreased considerably in the other two experimental conditions: Non-Communicative (NC) context (in which the agent did not look at the infant and her body was rotated 90° away from the infant); and in Non-Social (NS) context, where the demonstrator was not visible by the child, and performed the actions of the game behind a curtain. However, the percentage of errors rested high in both two groups: 57% in the former, 64% in the latter. Therefore, the novel adaptation of Piaget's experiment did not provide clearly the indirect evidence of the reasons why mistaken interpretation persisted. Infants appeared to search randomly the object under A or B containers. Although natural pedagogy's application does not manage alone to take into account the high percentage of errors in non-ostensive conditions, its explanatory strategy shows that alternative explanation of motor mirroring provided by Longo and Berenthal (2006) is wrong. Indeed, how Topál and colleagues illustrate, their results

> challenge recent proposals that the motor priming of the prepotent response can be induced by simply observing the manual hiding actions directed at location A, mediated by the mirror neuron system […], because the NC and OC contexts provided the same amount of visual (as well as motor) experience of the repeated manual hiding actions directed at container A (Topál et al. 2008, p. 1833).

Topál and colleagues proposed that A-not-B paradigm is interpreted by infants as «a kind of teaching session that conveys generalizable information about properties of the objects (*toys or containers*) for the infant to learn». This kind of incorrect

interpretation - established in virtue of the ostensive demonstration (A-trials) - «remains dominant during the B-trials, leading to the erroneous perseverative search responses» [my italics] (Ibidem). Therefore, Vorms (2012) is right when she claims that it is not explained exactly what kind of information is conveyed. Does information regard object functions and its features or its location? Although some hypothesises can be advanced in this sense, the A-not-B error task rests a developmental puzzle that natural pedagogy alone cannot explain completely, as Topál and colleagues acknowledge too. We assume that infants must understand the aim of the game, that is to retrieve the object. However, why should they do? Maybe, the structure of the events presented in that way does not allow to infer a clear teleological interpretation[16]. However, action interpretation framework remains the best explanatory strategy that can clear (any) doubt away about the potential interpretative oscillation between object representation and its spatial location. In particular, according to the pedagogical account, it could be the case that object representation is profoundly connected with and dependent on the action by which the object is presented. «Kind-defining properties are bound to the actions one should perform with them» (Vorms 2012, p. 540). In this sense, location might be a relevant property of the object. In other words, it is the entire action with the spatial relations between object and containers, and their characteristics to be interpreted as it were: «container A is 'for' storing the kind of objects being hidden» (Csibra&Gergely 2009, p. 152). Vorms reformulates this statement as following: «one should generally store this kind of objects under container A»[17] (Vorms 2012, p. 539).

Vorms' formulation properly emphasises the normative implication of the pedagogical learning. «Descriptive information about the function of artifacts – Vorms writes reporting Gergely's thought - involves a normative dimension about how one

[16] Pierre Jacob suggested Vorms to create a helping configuration of the experimental setting that would allow to separate the action interpretation from the difficulties tied to executive functions. For example, another test could measure through VOE paradigm «infants' surprise in front of another agent performing the A-not-B task (alternatively correctly and wrongly)» (Vorms 2012, p. 543).

[17] This means that «infants would rather generalize information about the object being hidden under a certain container (A)», rather than «information about the object being hidden under a container that stands in a certain location (A)» (Vorms 2012, p. 541).

should use it»[18] (Vorms 2012, p. 539). The normative dimension is the direct consequence of the cognitive biases that serves to fix, *without further evaluations*, the epistemic content acquired. The latter goes to settle in the backdrop of belief network on the basis of which children can make inferences about the worldly events, and evaluate the trustworthiness of future information (testimonies) transmitted in the most diversified contexts.

5.3 Nakao and Andrews' critique: around the notion of opacity

Instrumental actions and artefacts «tend to be opaque both in terms of their adaptive function (*teleological opacity*) and in terms of their modus operandi (*causal opacity*)» (Csibra&Gergely 2011, p. 1154). According to the natural pedagogy stance, infants and children assume that adults perform their actions rationally and efficiently to demonstrate novel and culturally relevant information under conditions of cognitive opacity (Gergely&Csibra 2005; Gergely et al. 2007). Natural pedagogy may guarantee a sufficiently stable transmission of complex (opaque) cultural behaviours across generations. Through its cognitive device, natural pedagogy system allows high fidelity imitation that worries the transmission of cumulative culture. Quoting Legare and Nielsen (2015, p. 688), «cumulative culture requires the high fidelity transmission of group-specific instrumental skills and social conventions to future generations». However, in a cumulative culture based society, besides imitation there is innovation, which «is necessary to ensure cultural and individual adoption to novel and changing ecological challenges over time» (Legare&Nielsen 2015, p. 689). Cultural technologies are not merely learned through imitation of others' behaviour. Rather, «imitation and innovation work in tandem to support cultural learning in children and facilitate our capacity for cumulative culture» (Legare&Nielsen 2015, p. 688). Such interaction between technological innovation and high fidelity social learning leads «to the development of increasingly complex technology over generations» (Moore 2016a, p. 35).

[18] Futó et al. (2010) and Butler&Markman (2012) pointed out this crucial aspect of natural pedagogy theory through their experiments discussed in chapter 3.

Natural pedagogy coupled with teleological reasoning enables infants and young children «to represent the structure of novel actions even when the causal relations between the method and the outcome are opaque and appear to violate the expectation for behavioural efficiency derived from the principle of rationality» (Legare&Nielsen 2015, p. 690). Király et al.'s study (2007) illustrates that infants are incline to imitate «the agent's teleologically opaque action only if the demonstration phase was run in an ostensive-communicative context» (Vorms 2012, p. 529).

Despite the early sophisticated capability to reason causally, «much of what children need to learn and interpret is not based on understanding physical causality and instead is based on social conventionality» (Legare&Nielsen 2015, p. 690). Conventional behaviours tend «to be associated with higher imitative fidelity than instrumental [behaviours]» (Legare&Nielsen 2015, p. 693). The findings presented by the two authors show «a lack of sophistication in tool innovation in young children and a reliance on being shown task solutions by adults» (Legare&Nielsen 2015, p. 696). For example, in Nielsen (2013) 4-year-olds were in front of a toy placed at the bottom of a Perspex tube. In order to retrieve it, children did have to add «water form a water bottle, causing the toy to float to the top». (Legare&Nielsen 2015, p. 695). Despite the availability of the bottle of water, few children spontaneously found the solution. The majority of them did reach the toy after the adult's demonstrative action, even when the method of the adult involved unnecessary passages, e.g. «pouring from the bottle into a cup before pouring from the cup into the tube» (Nielsen 2013, p. 44). According to Nielsen, «these results highlight the value of overimitation to children growing up in a world filled with objects whose operating mechanisms are often hidden or unclear».

6 Natural pedagogy and language according to Mattos and Hinzen

Mattos and Hinzen (2015, p. 2) put forward the hypothesis that «natural pedagogy is inherently integrated with language», or in other words that «the faculty of language and the capacity to learn from communication are intrinsically related» (Mattos&Hinzen 2015, p. 6). According to them, «there would be a single evolving system, and the prediction is that natural pedagogy and language will never dissociate» (Mattos&Hinzen 2015, p. 2). Obviously, communication is one of the main sources of

knowledge for humans, but it is also true that communication does not occur only through verbal channels. Furthermore, knowledge by communication is spread among animals. The authors' strong assumption is that *all of the human-specific forms of communication* have a linguistic structure. In particular, they suggest that the interpretation of «declarative gestures by infants reflect structural aspects of human language» (Mattos&Hinzen 2015, p. 2). As I hope to have clarified in the course of the dissertation, natural pedagogy theory is grounded on a crucial aspect of the Gricean theory that has been elaborated by the relevance theory under a cognitive point of view: the very public component of the communicative act triggers a huge referential expectation which enables infants to strongly metabolised the informative content in a social direction through the generalizability and universality biases.

In Mattos and Hinzen's perspective, natural pedagogy can transfer generic knowledge about kinds and trigger the generalisation bias in virtue of what the authors call the «emergence of proto-determiner phrases» (DP) (Mattos&Hinzen 2015, p. 7). According to them, the DP-stage enables young children to acquire «knowledge about kinds» (while *proto-sentence* ability allows «knowledge about facts») (Ibidem). As Prasada (2000, p. 67) claims, the crucial aspect of knowledge about kinds is that «[it is] not rendered false by the existence of instances that lack *the essential property*» [my italics]. This means, for example, that the specific case of an entity that lacks a characteristic element does not impair the entire category to which it belongs. Therefore, following Mattos and Hinzen's reasoning, the DP ability implements both generalizability assumption and the essentialist bias. The authors do not mention explicitly the essentialism, but the reference to Prasada clarifies this connection which I have already emphasised in a prior chapter (Ch. 3, §§ 7 and 7.1). However, the causal relation put forward by Mattos and Hinzen is too demanding. In my opinion, their intriguing suggestion argues more cautiously in favour of a common ground between the ability to conceptualise according to kinds and the assumption of generalizability. By postulating that the former is the prerogative for language acquisition, they conclude that natural pedagogy has a linguistic root. Instead, in a more parsimonious way, we can affirm that DP fosters and elicits objects-conceptualisation even in a pedagogical situation. To the same extent, as I have argued in the course of this dissertation, the initial form of infant mentalisation promotes the trigger of a universal informative

contents-attribution. Rather than a strong causal relationship between such variety of cognitive systems, I prefer to imagine the infant's mind as a network of autonomous cognitive systems that, by incorporating functions, cooperate or compete each other according to the contexts and developmental stages.

7 Limits of natural pedagogy

The ideal pedagogical relationship envisages a knowledgeable individual (the 'teacher') who transmits voluntarily, through an intentional communicative acts, information for the benefit of one (or more) naïve individuals (the 'learners'), with the intention of fostering learning in them. «The information provided by the knowledgeable individual is generalizable, or relevant to the identity of the group to which teacher and students belong, and could serve as a platform for future insight or innovation by others» (Moore 2016a, p. 41).

«Pedagogy promotes *efficient learning* but at a cost: children are less likely to perform potentially irrelevant actions but also *less likely to discover novel information*» (Bonawitz et al. 2011, p. 322) [my italics]. In this regard, I have to mention the experiment conducted by Pinkham and Jaswal (2011) who modelled the lightbox-experiment in order to investigate whether 18-month-old infants' prior experience with the magic box affected their imitation in pedagogical demonstration. They asked whether infants «would imitate an adult who used her head to illuminate a light-box if they had earlier discovered that the light could be illuminated with their hands» (Pinkham&Jaswal 2011, p. 535).

> In the Self-Discovery condition, infants had the opportunity to freely explore the light-box; all infants used their hands to activate the light-box at least once during this period. The experimenter then entered the room and, while providing explicit pedagogical cues, demonstrated illuminating the light-box using her forehead. In the Demonstration Only condition, infants just viewed the experimenter's demonstration. During a subsequent testing phase, infants in the Demonstration Only condition were more likely to use their foreheads to activate the light-box. Conversely, infants in the Self-Discovery condition were more likely to use their

hands, suggesting that efficiency can ‘‘trump’’ pedagogy in some observational learning contexts (Ibidem).

This important experiment has been used to criticise natural pedagogy model (Nakao&Andrews 2014; Heyes 2016) because it would suggest that «in infancy even rational imitation does not trump or overwrite individual learning in a way that would allow culturally accumulated wisdom to be passed down from one generation to the next without corruption» (Heyes 2016, p. 287-288). However, the advocates of natural pedagogy theory never claimed that natural pedagogy social learning device has the force, or the aim, to revise previous beliefs or practical knowledge gained in other ways or in another context. Moreover, even if infants recognise to be involved in an ostensive context they might not understand the real task, and otherwise, misinterpret the pedagogical condition as an opportunity to demonstrate their own knowledge.

6.1 Natural pedagogy and deep certainties

Stable knowledge emerges not only from the direct experience of the world, but also from indirect experience of testimony. Children learn by trusting and then believing the adult through the transmission of testimonies, and the knowledge achieved forms a system of convictions that constitute common ground of beliefs. The concept of common ground is dramatically important for navigating properly in the social environment, understanding others’ behaviour, and sustaining adequately conversations. For example, Stalnaker (2002), who has systematically developed the pragmatic nature of ‘common ground’ starting from Grice’s idea of presupposition, claims that the full comprehension of conversations depends on the presence of shared beliefs and the acceptance of some assumptions even if they are not believed at all[19].

[19] Stalnaker’s notion of presupposition predicts that: 1) the relationship of presupposition is not a semantic relationship between utterances, but a relationship between propositions and speakers: having presuppositions means having an attitude towards a proposition, namely the content of an utterance; 2) the presuppositions do not interfere with the evaluation of the truth or falsity of the propositions but with their *appropriateness* relatively to the context; 3) additionally, since presuppositions are assumed to be the background of speakers, they are not necessarily connected to the use of linguistic expressions (Domaneschi&Penco 2017).

In conclusion, I hope to be clear the crucial role I suppose the natural pedagogy system plays in the construction of a stable network of beliefs relatively to some domains. I have tried to border the contextual application of natural pedagogy that represents only one of social learning strategies employed by "teachers" and pupils. Natural pedagogy system, based on child-directed interactions and in cooperation with other cognitive system like early mindreading system, can represent one of the most efficient adaptive strategies to create firmly that deep «nest of propositions» which - Wittgenstein (1949/1969, § 102; § 225) wrote in *On Certainty*[20] - build cultural shared beliefs structure to be relied on and followed. We may claim that through natural pedagogy children can learn a host of things belonging to cultural knowledge domain, and they learn to act according to such beliefs, to such know-how. In addition, following once again Wittgenstein, this body of knowledge is not believed by a single individual, but the single subject believes that others believe it too. «I am firmly convinced that others believe, believe they know, that all that [body of knowledge] is in fact so» (1949/1969, § 289). This can be considered as an alternative formulation of universality assumption predicted by natural pedagogy. By the emergence of such particular harmony with the last Wittgenstein we can express with his words a concept already emerged during this dissertation:

> «Bit by bit [children form] a system of what is believed, and in that system some things stand unshakeably fast and some are more or less liable to shift. What stands fast does so, not because it is intrinsically obvious or convincing; it is rather held fast by what lies around it» (Wittgenstein 1949/1969, § 144).

Children did not get that nest of propositions (which depicts to some extent their own picture of the world) by satisfying their selves of its correctness, nor do they have it because they are fully satisfied of its correctness, but because «it is the inherited background» (Wittgenstein 1949/1969, § 94). In conclusion, borrowing another powerful Wittgenstein metaphor, we can compare such kind of beliefs-network with a riverbed that «consists partly of hard rock, subject to no alteration or only to an

[20] (Wittgenstein 1949/1969) «§ 102 […] Yet my convictions do form a system, a structure» «§225. What I hold fast to is not one proposition but a nest of propositions».

imperceptible one, partly of sand, which now in one place now in another gets washed away, or deposited» (Wittgenstein 1949/1969, § 94).

CHAPTER 7.

Beyond the limits.
Social Biofeedback and Natural Pedagogy theories

1. Introduction

In this final chapter, I present the hypothesis advanced by Gergely (together with other psychologists) about the supposed cooperation between the natural pedagogy model and the social biofeedback model, that takes into account the infant's internalisation process of *contingent "marked" emotions-mirroring displays*. Such affective mirroring manifestations involve the generation of the infant's second-order representations of primary non-conscious affective Self-states. Such basic affective states are initially perceived by infants either positively or negatively. Instead, the second-order representations become cognitively accessible and allow the progressive conquest of subject's introspective awareness able to discern anger, for example, from an indistinct negative arousal.

Despite agreeing with the hypothesis concerning the social construction of a subjective sense of Self - that is not the topic discussed here rather the assumption of the argumentation -, I reject any involvement of natural pedagogy with social biofeedback, and thus, with the construction of the child's inner emotional Self. I will show that Gergely's theoretical proposal is based on the equivalence between marked affect-mirroring displays and infant-directed cues of ostensive communication intended as referential knowledge manifestations, where 'knowledge' stands for the variety of emotions. In a few words, according to Gergely's hypothesis, the grounding elements of the natural pedagogy system would make possible also the social construction of the infant's inner emotional Self. Such claim entails that inner emotions, like cultural knowledge, are *taught* by adults to infants through social interactions based on the referential-expectation power of ostensive communication. By rejecting this kind of cooperation, I am not going to deny that infants learn to regulate and express their own emotions via continuous social interactions with attachment figures. However, in my opinion, this pragmatic learning does not pass through an explicit *pedagogical* teaching. In the previous chapter, I have clarified that the presence of ostensive cues is not fully

sufficient to establish pedagogical conditions. Even if we assume that the 'markedness' of emotion displays is a pragmatic form of ostensive communication, the automatic involvement of pedagogical stance is not guaranteed or implied by sharing the referential nature of ostensive cues. In fact, first of all, pedagogical stance requires a high degree of intentionality: *the intention of teaching*. Such manifested intention is not so evident during parental affect mirroring consisting, instead, in a dialogical *negotiation* (Grifftiths&Scarantino 2009) that is not established in the name of a declared *teaching* intention. Furthermore, as I sketched in the previous chapter, the pedagogical transfer of cultural knowledge, compared with the transmission of testimony, implies propositional forms for the conveyed contents that cannot be also applied to emotions so easily. In fact, we should admit that primary emotions are not conceptual, while second-order representations have propositional form only in force of a *pretence strategy* fostered by *referential expectations*. Such scenario that I will discuss below would deserve to be tackled deeply, and I hope that further research will do it.

Finally, I conclude that social biofeedback and natural pedagogy are not dependent on or hierarchically connected to each other. On the contrary, they are autonomous cognitive systems which share (and are based on) infants' great sensitivity to interpret the referential nature of ostensive communication.

1.1 The two assumptions of social biofeedback theory: lack of emotional introspection and the social construction of the infant's inner emotional Self

Against Cartesian view that considers the recognition of subjective internal states as transparent, accessible, immediate and guaranteed even from infancy (see, for instance, the influential «primary intersubjectivity» model[1]), in developmental psychology a social constructivist perspective has emerged. According to constructivist view, the individual rich emotional internal states are the result of a long social apprenticeship due to repeated interactive experiences of infants with their caregivers and specifically with attachment figures (for a review see Caruana 2017). In a few

[1] Cfr. Trevarthen (1994), and Trevarthen&Aitken (2001). The primary intersubjectivity model attempts to find in MNs neurobiological foundations.

words, the subjective sense of affective states has social environmental origins and it is not given from birth (Fonagy et al. 2007, p. 289).

The social biofeedback model assumes that human infants initially show a primary bias «to construct representations mainly based on exteroceptive stimulation [and] leads to the construction of discrete emotions» (Di Francesco et al. 2016, p. 119). In this view, (discrete) emotions emerge from raw precursors, i.e., the basic emotions delineated by Ekman (1992), which are «the most elemental among discrete emotions […] biologically based and pancultural packages of short-term, coordinated and automated responses to events in the environment» (Di Francesco et al. 2016, p. 120). Generic categories of the basic emotions like fear, anger, joy, pleasure, disgust are universally recognised and shared by human beings. Basic emotions are probably the evolutionary product of adaptive automatic systems that are present also in many species of non-human animals including primates (Gergely&Unoka 2008a; Panksepp 1998; Suomi 1999). The predetermined behavioural response triggered by basic emotions are termed «affect programs» (Ekman 1992). From affect programs, structural procedures emerge including a variety of components as physiological changes (somatic component), facial and vocal expressions (motor component), and intentional acts (motivational component). Therefore, infants would be endowed with few emotion categories representing the indistinct inner magma of sensations that little by little must become separate and distinct (Griffiths&Scarantino 2009; Scarantino 2012; Widen 2013) during a normal ontogenetic development. The emotion states are «not grouped together categorically in such a manner that they could be perceptually accessed as a distinctive emotion state» (Gergely&Watson 1999, p. 110; see also Gergely&Unoka 2008a, p. 62; Di Francesco et al. 2016, pp. 118-119). The best definition for depicting the initial emotional jumble has been provided by Russell (2003) with the notion of «*Core Affect*», that consists in a «neurophysiological state that is consciously accessible as a simple, non-reflective feeling that is an integral blend of hedonic (pleasure-displeasure) and arousal (sleepy-activated) values» (Russell 2003, p. 147; for a review see Marraffa&Viola 2017; Caruana&Viola 2018).

In order to reconcile and harmonise such innatist aspect with the rest of constructivist features of social biofeedback theory, Marraffa&Meini (2016), and Meini (2017) suggest to replace Ekman's approach with the new theory of basic emotions

(New BET) provided by Scarantino and Griffiths (2011; Scarantino 2015). New BET, in fact, keeps the modularist-innatist approach but it makes room for constructivist integrations. «The raw experience produced by the modules of emotions acquire progressively a distinct form in an interpersonal context» [my translation] (Meini 2017, p. 117). According to Griffiths and Scarantino (2009, p. 447), «emotion is shaped by the ongoing process of negotiation». In other words, emotions are forms of «engagement with the social environment that involves a dynamic process of negotiation mediated by reciprocal feedback [provided by emotional signals] between *emoter* and *interactants*» [my italics] (Griffiths&Scarantino 2009, p. 446).

Affective communicative expressions are mutually influenced and bidirectionally regulated within the caregiver-infant dyad by two-three months of age when infants show active social expressions during face-to-face interactions. Caregivers' facial and vocal displays in response to infants' displays are not randomly but emotion-specific. Infants actively attend to the so called (in developmental literature) *contingent affect-mirroring interactions* (see e.g., Murray&Trevarthen 1985). Murray and colleagues (2016, p. 7) emphasise that «the form of parental response is critical, indicating the experience-expectant nature of the development process». They show that «the preparedness of the infant's perceptual system to attend to particular social stimuli is matched by the propensity of parents to deploy specific responses to different infant behaviours». This allows young infants «to capitalize on relatively limited exposure to specific parental behaviours, in order to develop important social capacities» (Murray et al. 2007, p. 1,7).

2. The contingency-detection device

The social biofeedback model applies contingency perception mechanism to parental affect-mirroring. As it is known, from a very early age, infants are highly and positively sensitive to causal relations between stimulus events and their physical reactions. The detection of causal control over contingently physical events provoke positive arousal in 2-month-olds who, for example, visibly start to smile and coo. Gergely and Watson (1999) proposed «the existence of an innate *contingency-detection module* that analyses the probability structure of contingent relations between responses

and stimulus events» [my italics] (Fonagy et al 2002, p. 163). The contingent response-stimulus events relation relies on a conditional probability structure which operates in virtue of two independent indexes: the *index of necessity* and the *index of sufficiency* (Watson 1994). The former looks «backward in time, monitoring the relative likelihood that a given stimulus event was preceded by a given response». The latter, instead, «looks forward in time, registering the conditional probability of an upcoming stimulus event as a function of an emitted response» (Fonagy et al. 2002, p. 163). According to Gergely and Watson (1996; 1999; Gergely et al. 2010), the innate contingency perception mechanism enables infants to estimate the conditional probability of three types of *causal contingent relation* between their own actions and effects in the surrounding environment. The three relations are: 1) temporal contingency between two events, 2) spatial similarity and 3) correspondence of relative intensity.

Surprisingly, Watson (1994) found that after 3 months of age, infants cease to be interested in those events where the relation between stimulus and response is perfectly contingent, while they tend to look at (and look for) those situation where the level of contingency remains high but not perfect. Therefore, Watson (1994) hypothesised that the initial goal of the contingency-detection module is to seek perfectly response-contingent stimulation in order to form a «*primary representation of the bodily Self*» (Gergely 2002, p. 28) *as a distinct entity in the environment* in which infants can exercise a perfect control. However, after three months of age the contingency analyser switches the preference to high-but-imperfect value of contingencies that orients the infant towards the social parental environment (Gergely et al. 2010). Therefore, the contingency-detection device provides information about emotional interactions that infants engage with caregivers through two kinds of basic emotional representations: infant's own causal emotion displays towards the caretaker, and caretaker's causal emotion displays towards the infant (Gergely&Unoka 2008a). This interpretative capacity of causality permits the beginning of dialogic (pragmatic) relation called properly affective-parental mirroring interaction, that starts to occur, thus, before the first birthday, and serves the function of anticipating and modulating the emotional consequences of particular types of attachment interactions (Gergely&Unoka 2008a).

3. Markedness of emotion expressions and its implications

Gergely (2007b, p. 63) suggested to include among the domains of generic knowledge that can be pedagogically transferred also

> «relevant *knowledge about the existence and culturally shared dispositional contents of specific categorical emotion states of the infant's constitutional self* that are initially made accessible to the infant through the caregiver's pedagogical communicative interactions involving ostensively "marked" contingent mirroring displays».

Firstly, Gergely and Unoka (2008a; 2008b), and subsequently Judit Futó (2010) argued in favour of this line of thought. Specifically, Futó affirmed that natural pedagogy system

> «can also be employed in the domain of emotion socialization to identify […] and transfer relevant knowledge about those categorical emotions to the infant that are culturally universal and shared among humans. In this view, early infant-caregiver affective interactions (involving ostensively cued 'marked' forms of contingent emotion-mirroring) constitute a special case of pedagogical knowledge transfer whereby sensitive caregivers establish second-order representation in infants that identify and encode categorical emotions» (Futó 2010, p. 19).

The marked forms of contingent emotion-mirroring, or 'markedness', are an exaggerated version of motor execution of standard expressive and responsive caretaker's display. Such exaggerated expressive forms help infants to identify communicative manifestations. In other words, motherese speech, motionese, and all types of infant-direct communicative modified manifestations can be considered – according to Futó – as forms of «marked contingent affect-mirroring interactions» that pedagogically would allow parents to transfer «culturally universal categorical emotions that are shared among humans» (Futó 2010, p. 58). According to Judit Futó (2010, p. 6), «Marking, *as ostensive cue* informs the addressee about the communicative and referential *goal* of communicator, and marked forms of referential *knowledge*

manifestations, emphasize new and relevant aspects of the communicated information» [my italics]. However, some troubles rise immediately. Firstly, the question about *goal*. If Futó is right, thus, within the mirroring dynamic described by social biofeedback, the markedness must be the expression of a teaching intention. This must imply coherently that the emotion-mirroring is lived by caregivers as a teaching practise and is interpreted by infants as a learning occasion. If in the parental mirroring the markedness is an ostensive cue, then the markedness is at the same time the content of the information transmitted. In fact, the very markedness is communicating infants not only that there is something to be communicated, but also it is telling that the content of the message is the exaggerated emotion itself.

Accepting that markedness is an ostensive cue entails the recognition of its peculiar function, which consists in signalling that the caretaker's display is not referring to caretaker's own internal state or contingent dispositional reaction, but it is referring to «infant's current internal dispositional state» (Futó 2010, p. 72). Therefore, markedness inhibits «the attribution of the emotion to the caregiver as her "real" feeling» (Gergely 2007b, p. 67). In other words, markedeness has the function to referentially *decouple* the representation of emotion-expression (Gergely 2007b; Gergely&Watson 1996; 1999; Gergely et al. 2010). Specifically, Gergely wrote that «in the interpretation of the marked affect-display the referential connection between the emotion expression and the corresponding dispositional state of the agent producing the display will be suspended: the perceived emotion display will be "decoupled" from its referent» (Fonagy et al. 2002, p. 178; Gergely 2007b). In this way, the exaggerated expression «will be represented as 'not being about' the caregiver's actual emotion state» (Di Francesco et al. 2016, p. 122). Since the caretaker looks insistently at and rests oriented towards the infant while producing marked emotion mirroring displays,

> «the infant's attention will be directed towards her own face and body, i.e., her own physical self as the spatial locus of the referent entity that the caregiver's attention orienting referent identification cues indicate and to which the 'marked'—and 'decoupled'—affect display should be referentially 'anchored'» (Gergely 2007b, pp. 67-68).

The referential effect of decoupling is depicted, thus, as an *anchoring*, a suspension of actual consequences which leads infants to establish *second-order representations* to associate with the infant's primary basic emotion states. Therefore, «the infant will set up *separate representations* for the affect-mirroring displays» [my italics] (Di Francesco et al. 2016, p. 122). Gergely specifies that:

> [the] process of *referential anchoring* is determined by the high degree of contingent relation between the parent's affect-reflecting display and the infant's emotion-expressive behavior. The infant's contingency-detection system will register the temporal contingency and cross-modal similarity of pattern between the parent's expression and his own on-going affective behavior. The perception of this contingent relation will provide the basis for the referential interpretation and grounding of the decoupled emotion display. As a result, the infant will referentially anchor the marked mirroring stimulus as expressing his own self-state» (Fonagy et al. 2002, pp. 178-179).

To sum up with Gergely's words, ostensive marked affect-mirroring procedures «"*teach*" the infant about the existence of her internal subjective emotion states, [and] lead to the *internalization* of the caregiver's "marked" mirroring displays as second-order representations [which] form the basis of a subjective sense of awareness of *internal self states*» (Gergely 2007b, p. 68).

4. Pretend component

The notions of «referential decoupling» and «referential anchoring» were first adopted by Alan Leslie (1987; 1994) to characterise the representational properties of communicative manifestations performed in pretend play. According to Leslie (1987), in pretence the truth reference of a first order expression (like for example 'this is a banana') is suspended and the same referent appears in a second-order representation, i.e., in metarepresentational context where it can be manipulated. In this way, simultaneously 'the banana is not a banana', and it can become a telephone without any contradiction. Therefore, the simultaneous occurrence of two representations engenders the problem of what Leslie called «*representational abuse*» (Leslie 1987, p. 415). For

Leslie (1994), «the decoupled representation is the object of another computational process» performed by an individual (Meini&Voltolini 2010, p. 38).

While Perner (1991) holds that pretence has not metarepresentational features - because a metarepresentation is a representation of a representation, «that is as something which has an interpretation and not as something which at most is a mere (syntactic) vehicle of information» - Leslie, on the contrary, claims that metarepresentation is «just an informational relation between an agent and two representations» (Meini&Voltolini 2010, p. 39; see also Friedman&Leslie 2007). On their turn, Meini and Voltolini (2010) acknowledge too that pretence involves metarepresentations (even if in a weaker sense than Leslie thought).

An intriguing proposal in line with Leslie's account has been advanced by Gómez (2008), who studied the so called 'linguistic apes', or 'enculturated apes', and compared their behaviours with infants' pretend play. Gómez proposed «a model according to which pretence is supported by a mechanism capable of computing *intentional relations with nonexisting objects* or properties (*Intentional non-existence*), as opposed to mechanisms computing *intentional relations with existing*, although not necessarily currently perceived, objects (*Intentional availability*)» [my italics] (Gómez 2008, p. 586). Connecting referential intentionality with pretence may be adequate for social biofeedback model. Nevertheless, further researches have to clarify the developmental timing of pretending phenomena, because the onset of pretending play has been attested around 18 months, even if the comprehension of pretence is anticipated and attested around 15 months of age (e.g., Bosco et al. 2006). However, so far, the attested emergence of pretence seems too late in respect of what social biofeedback model requires.

In this regard, I repute very important the longitudinal study conducted by the researchers Futó, Bátki, Koós, Fonagy and Gergely (2004). They tested «the developmental relation between contingent maternal mirroring and "markedness" of contingent maternal reactions in a group of 12-month-old infants, on the one hand, and different aspects of pretence competence of the same children at 2.5 years of age, on the other» (for a review Gergely 2007b, p. 73; Futó 2010, pp. 69-80). The developmental relation was measured by adopting a modified version of Still-Face paradigm (Tronick

et al. 1978), the so called Mirror Interaction Situation (MIS) (Koós&Gergely 2001). The researchers designed three phases in which

> «mother and infant were seated next to each other in front of a one-way mirror. They were separated by an occlusion screen that prevented them from physically contacting each other: however, they were free to interact facially and vocally with each other's' mirror image. […] After a first phase of free interaction, the mother was instructed to put on a motionless, neutral "still-face", looking at but not reacting to her infant» (Gergely 2007b, p. 73).

The second phase provoked mild stress in the infant. However, the deprivation of maternal reactivity was immediately followed by a third phase during which the mother could interact again with her child. Once recorded these data, the researchers followed up a sample of toddlers between 2 and 3 years of age in order to test during the individual growth the capacity to develop pretence skills. They modified and enriched a previous version of pretence task (Harris&Kavanaugh 1993) «to measure representational aspects of pretence competence along the lines of Leslie's (1987) analysis of the metarepresentational structure and representational operations implied by understanding and producing pretend play» (Gergely 2007b, p. 73). In fact, Futó and colleagues (2004) wanted to compare both representational pretence competences and the capacity to elaborate pretence spontaneously and creatively too. What did they found? For what concerns 'markedness', during the first and third phases of MIS, a high degree of 'markedness' of contingent maternal references at 12 months of age predicted «both higher overall representational pretence competence scores and higher scores on spontaneous, adequate, and creative extensions in the use of pretence at 2.5 years of age» (Gergely 2007b, p. 74; Futó et al. 2004; Futó 2010). For what concerns the contingent maternal reactivity, during the first and third phases of MIS, a high degree of affect parental mirroring at 12 months «predicted high scores on spontaneous, adequate, and creative extensions of pretence performance at 2.5 years» (Gergely 2007b, p. 75). On the contrary, the low tolerance to deprivation of maternal reactivity at 12 months corresponded to «low pretence performance in open-ended pretence situations involving separation or physical injury at 2.5 years of age» (*Ibidem*).

5. What is shared and what is not by social biofeedback and natural pedagogy

In conclusion, I deem that social biofeedback mechanism encompasses a variety of complex devices (contingency-detection module, referential interpretations of mirroring displays, decoupling metarepresentational capacities), and interacts properly with variable dynamics of attachment relations through experiences and learning strategies. In line with Gergely (2007b) we can affirm that in the ontogenetic developmental trajectory

> «by *1 year of age* securely attached infants have already learned—as a result of their experiences with the "marked" contingent mirroring reactions of their sensitive, infant-attuned caregivers—to direct *at least part of their available attentional resources in an introspective direction*, actively monitoring the dynamic changes of their internal self states as well as the proprioceptive cues produced by their own emotion-expressive behaviours» [my italics] (Gergely 2007b, p. 78).

Natural pedagogy system can be excluded from this developmental framework without impairing or taking any functional component away. Indeed, natural pedagogy is involved (simultaneously) in other crucial tasks relying on some shared cognitive domains. The kind of relationships established with (probably) the same caregivers (as it is easy to guess) is different for what concerns the degree of intentionality. The distinction between the two contexts, *mirroring* and *teaching*, allows avoiding the erroneous equivalence between ostensive context and the trigger of pedagogical stance.

Although social biofeedback and natural pedagogy theories share social and motivational dimensions, the degree of explicit intentionality is different and discriminant in the two cognitive systems. I think that social bio-feedback shapes more frequent, unconscious, *fragile* dialogic relations that are dependent on caregiver's subjective emotional dispositions rather than on the natural tendency to teach explicitly knowledge contents about social external reality. In the latter, in fact, the attitude is more defined and the intentions are more explicit although in both cases ostensive manifestations serve the same end. The manifestations of such intentionality are conveyed through ostensive signals that can be flexibly used in other contexts and not

always with the aim of teaching something. The case of motherese and motionese is the best evidence. The adults' infant-direct attention expressed by exaggerated and modified vocalisations, smiles and other nice approaches does not always entail a particular communicative relation aimed to transfer knowledge. Natural pedagogy and social biofeedback, thus, do not share the same degree and form of intentionality.

The object centred-perspective that characterises the natural pedagogy model cannot become in the dyadic attachment relation a sophisticated self-centred device for naïve subjects interpreting their own inner emotional states. Rather, what the two systems share, pragmatically, is limited to the utilisation of the full referential power of ostensive signalling.

Moreover, the construction of distinctive emotions categories depicted so efficiently by the notion of *social negotiation of emotions*, cannot be considered as a special *achievement mediated by conceptual thought*. Emotions are primarily - Griffiths and Scarantino (2009, p. 439) claim - «effective goal-oriented responses». It would be difficult to explain their contents in terms of propositional attitudes, rather «emotional content has a fundamentally pragmatic dimension, in the sense that the environment is represented in terms of what it affords to the emoter in the way of skillful engagement with it» (Scarantino 2009, p. 441). However, the second-order representation of basic emotions, allowed by early pretence competence, probably would open the way to the possibility of precocious conceptual construction of emotions, as Gergely and Unoka (2008a) seem to suggest. In fact, I interpret in this sense their following claim:

> «with the developmental construction of cognitively accessible second-order representations of internal self states, *the proper domain of the human mindreading* becomes ontogenetically extended to include in its actual domain the mind of one's own self as well» [my italics] (Gergely&Unoka 2008a, p. 74; 2008b)[2].

This introduces a conceptual construction of emotions that could be in line with the conceptual act theory (CAT) proposed by Lisa F. Barrett (2006a; Barrett et al.

[2] They adopt the terminology of "proper domain of mindreading", because in their view mentalization is «originally restricted to inferring and representing the causal intentional mental states of other minds only» (Di Francesco et al. 2016, p. 127).

2015), who claimed that «discrete emotions emerge from a conceptual analysis of core affect […]. Specifically, the experience of feeling an emotion, or the experience of *seeing emotion in another person, occurs when conceptual knowledge about emotion is brought to bear to categorize a momentary state of core affect*» [my italics] (Barrett 2006b, p. 49).

For Barrett, this kind of categorisation is supported by language rather than pragmatic instructions taught by adults, while Gergely and Unoka (2008a) prefer to take into account the mentalization, intended as that cognitive process able to *predict* the social consequences of emotional states expressions. Furthermore – they claim – mentalization helps humans also *to predict which mental states and what actions can trigger in others the automatic Self responses induced by emotions*. Therefore, which is the task of natural pedagogy? Advocates of the potential cooperation between social biofeedback and natural pedagogy hold that the contingency-detection device, social biofeedback, and natural pedagogy play «a crucial role in establishing primary self–other affective relationship representations that capture the characteristic causal structure of contingent reactivity of early attachment relationships» (Gergely&Unoka 2008/2013[3]). On the contrary, I defend a more circumscribed application of natural pedagogy model. I argue that social biofeedback does not need natural pedagogy because the former uses other cognitive processes as contingency-detection, pretence representations combined with the capacity to interpret ostensive cues that are not employed pedagogically *as such*. Last but not least, social biofeedback predicts the intervention of mentalization.

It is evident that the 'conquest' of fine emotion categories is a hard path where the final goal is not ontogenetically guaranteed and depends largely on the subjective behaviours of caregivers. Thus, the social negotiation of emotions is overexposed to the constitutive fragility of the mirroring dialogic relation. In this regard, Fonagy, Gergely, Jurist and Target (2002, pp. 9-10) claimed that

> «children whose parents provide more affect-congruent contingent, and appropriately marked, mirroring displays facilitate this decoupling. In contrast, the

[3] Originally published in 2008, I refer to the subsequent on-line edition (2013) available to DOI: 10.1093/med/9780198569183.003.0011.

> displays of parents who, because of their own difficulties with emotion regulation, are readily overwhelmed by the infant's negative affect and produce a realistic *unmarked emotion expression* disrupt the development of affect regulation. A major opportunity for learning about the difference between representational and actual mental states is lost. […] *Affect-mirroring can take pathological pathways*, because the caregiver is overwhelmed by the negative affect generated in response to the infant's reaction and presents an overly realistic emotionally arousing display. This undermines not only the infant's possibility of creating a secondary representation, but also the sense of a boundary between self and other…» [my italics].

6. Psychopathological implications

From this quotation, we can ascertain, on the contrary, that the constitutive fragility of dialogic mirroring interactions cannot be confused with a sort of *pathology of teaching*. The affect-mirroring can be impaired because adults *are not able* to provide the necessary care, and not because there is an intentional misunderstanding of an alleged pedagogical message. The notion of «unmarked emotion expression» used by Fonagy and colleagues (2002) cannot be referred to the outcome of adult's deliberate and "bad" intention aimed to mislead and undermine the constitution of infant's inner affective Self.

The lack, or the impairment, of affective mirroring may imply psychological instability and serious mental disorders, as Borderline Personality Disorder (BPD), that can be considered the disorder par excellence able to show how the absence of «the positive effects associated with the capacity to benefit from the social environment» can generate insecure attachment that in turn undermines mental health (Fonagy et al. 2015, p. 575). Insecure attachment is associated with personality disorder as an extensive body of psychopathological literature attests. Furthermore, early disruptions of mother–infant communication predict «not only attachment disorganization but, in some carefully conducted longitudinal studies, the likely emergence of personality pathology» (Fonagy et al. 2015, p. 587; for a review see Fonagy&Luyten 2016).

In this regard, Fonagy and colleagues (2015) proposed «a developmental framework that conceptualises BPD in terms of a specific underlying vulnerability

[defined] as the impairment of epistemic trust» (*Ibid.*, p. 576). In their view, «attachment may be seen as part of a mechanism *of deferential knowledge transmission* that has evolved to create a kind of *epistemic connection* between learners and teachers» [my italics] (Ibid., p. 584). Although secure attachment is not necessary for generating epistemic trust, it is sufficient to create «a sense of epistemic trust» (*Ibidem*). Thanks to an intriguing finding conducted by Corriveau and colleagues (2009), Fonagy and cooperators emphasise that children with disorganised attachment tend «to mistrust both information from their own experience and the attachment figure's (or the stranger's views)», as if these children are not able to select which informative source is reliable, and which message is trustworthy (Fonagy et al. 2015, p. 588). Fonagy and colleagues suggest that this dilemma is due to «a potentially interminable epistemic search [that] may generate a state of epistemic *hypervigilance* where lack of trust is generalized to any communication» (*Ibidem*). In a few words, «the *destruction of trust* in social knowledge» is interpreted as «*the key mechanism in pathological personality development*» (*Ibid.*, p. 589).

In my opinion, here, the crux is not the pragmatic implication that links trust to social knowledge and, thus, to communicative transfer of knowledge (see Chapter 6 for more details about this topic). Rather it concerns the notion of infant's *epistemic demanding* hierarchically connected with attachment dynamics. Fonagy and colleagues' (2015) account rests on a strong theoretical assumption whereby the epistemic principle, that regulates the innate research of relevant information concerning the surrounding world, should underlie (not only natural pedagogy but also) the developmental trajectory leading to the construction of an emotional Self. This implies that the child tends to (learn to) navigate the social world in the same way that she needs to navigate her own inner world. Therefore, the epistemic principle should subordinate and regulate the human (and maybe not only human) search and need for care, security, comfort, consolation, affection, recognition which are normally provided by caregivers.

In the last decades, cognitive science has illustrated the strong interconnection between emotions, cognition and agency. Nevertheless, on the basis of the available experimental researches and given the critical states or the limitations of theoretical models such as social biofeedback and natural pedagogy, I think it is still premature and theoretically reckless providing an explanatory model of the origins of emotional

introspection based on the innate human tendency to acquire (cultural) knowledge. Finally, the weakness of such approach lies in sketching the equivalence between cultural knowledge and emotions-mirroring interaction only on the base of *communicative modalities of transmission*, without investigating the *form* and the *nature* of the conveyed contents.

CHAPTER 8.

Conclusions

The present dissertation has investigated one aspect of the natural pedagogy theory proposed by Gergely Csibra and György Gergely. They provide a model of social learning that is based on direct communicative ostensive relation, and it is aimed to the transfer of generic cultural knowledge.

The pedagogical transmission of information origins from an explicit manifestation of teaching made by knowledgeable adults, that can be fast and efficient in virtue of a natural inclination possessed by infants to seek information and to decode signals of ostensive communication. In other words, knowledgeable individuals are naturally inclined to provide manifestly their cultural baggage to naïve conspecifics, who are equipped to receive and assimilate the information transferred by *searching and attending* to such (communicative) manifestations (Gergely&Csibra 2006). Gergely (2007a, p. 173) defined natural pedagogy as «a *relevance-guided social communicative learning* device of mutual design» [my italics] able to guarantee an effective transfer of a relevant cultural knowledge through multimodal ostensive communication.

From birth, human beings are immerged in a world of social norms that regulate behaviour and involve the use of objects, whose functions are not immediately cognitively transparent. Gergely and Csibra argue that during hominid evolution natural pedagogy has been selected to ensure fast and efficient acquiring and intergenerational transfer of cultural knowledge overcoming the hard social environmental conditions of «cognitive opacity» of human cultural forms (Gergely&Csibra 2005; 2006; Csibra&Gergely 2006; Gergely 2007a; Gergely 2013).

Natural pedagogy combines adult/infant cooperation into the form of a natural tendency to inform and to receive information. By definition, communication is intended as «a cooperative act that occurs in the context of joint attention between at least two intentional agents» (Vouloumanos&Onishi 2013[1]). Such definition is perfectly consistent with natural pedagogy prescriptions. Against Tauzin (2017), I posit that a

[1] I refer to the on-line version. DOI 10.1093/acprof:oso/9780199988341.003.0010.

minimal degree of cooperation is assumed by natural pedagogy theory, and it is based on joint action, i.e. sharing attention through ostensive signals towards a referent. The two poles of the communicative relationship are held together by a primary form of cooperation that is in line with the Gricean principle of cooperation, as I have underlined in Chapter 1, where (§ 6) I emphasise also the profound debt of natural pedagogy to the relevance theory (Sperber&Wilson 1986/1995; Wilson&Sperber 2002).

1. Epistemic trust as the basic assumptions of natural pedagogy theory

As David Sloan Wilson wrote, without *cooperation* and *trust* «the ability of group members to learn from each other in a cumulative fashion and to transmit their learned behaviors across generations» becomes strongly limited (Wilson 2013[2]).

Within the theoretical framework of natural pedagogy, epistemic trust is the fundamental assumption allowing a faster achievement of the manifested information, guaranteeing the young learner to fix this knowledge content. Epistemic trust is a crucial component of pedagogical stance because it allows the *deferential attitude* (well described by Recanati (1997)) that predisposes the juvenile audience (in this case) to consider others as benevolent and reliable sources of cultural information, activating the *presumption of relevance* about manifested knowledge contents. In this way, cultural contents can be transmitted and accepted in virtue of the authority of the source. The reliability of the source, conquered in virtue of the epistemic trust, represents likely an assurance for the benefit the information transmitted (say *p*), namely, an assurance of the validity and the importance of *p* from the recipient's perspective.

The acceptance of the content, that I treat as a testimony (see Chapter 6), is given by the typical interpretation of asserted and transmitted message. An implicit acceptance agreement is 'ratified' between counterparts, permitting in this way not only the mere communicative transmission but also the transformation of conveyed information into (*propositional*) knowledge. This informational baggage, in turn, constitutes those *common ground* beliefs that serve both to evaluate (i.e. provide justification for) further testimonies, and to support those pragmatic presuppositions useful for sustaining social communicative interchanges. In natural pedagogy, trustworthiness of testimony is

[2] I refer to the on-line version.

combined with reliability of knowledge source. Therefore, at the beginning of our belief foundational process we have *acceptance* of testimony based on an act of *trust*. Without at least some measure of trust, «one might not acquire testimony-based belief at all» (Audi 2015, p. 256). The very fact that the infant has formed beliefs, i.e., the fact that representation of informative content is stored in his/her belief box, constitutes a sufficient justification for persisting to hold it, and ascribe it to others.

However, as suggested by Gergely and colleagues (2007), development mitigates the strength of such deferential attitude induced by pedagogical cues. In fact, deferential attitude has to change in favour of what Sperber and colleagues (2010) have defined *epistemic vigilance*, «that assess the quality of incoming information and the trustworthiness of the individual who dispenses it» (Mazzarella 2016, p. 183). As already suggested by Koenig and Harris (2007) and confirmed by Sperber et al. (2010), such vigilance is both directed to the content of knowledge and to the communicator. The increase of epistemic vigilance inhibits the fast and low-level cognitive processes that allow, in virtue of the eliciting presence of ostensive cues, to recognise communicative intentions and trigger referential expectations. I suggest (Chapter 6) that the increasing reasoning capacities and the maturation of evaluating competences of the source reliability inhibit natural pedagogy system and circumscribe the applications of such social learning device throughout development.

Trust is gained by the informative sources through the niceness of her/his communicative approach. The advocates of natural pedagogy ground their theory on the great sensitivity shown by infants to capture and decode those ostensive signals which have the main task to indicate clearly that there is something important that can be learnt.

2. Early comprehension of the referential power of ostensive cues

For a full-blown ostensive communication it is necessary that infants understand not only the referent targeted by the gaze direction but, first, that they are the addresses of ostensive signals. According to Hoehl et al. (2008), by 4-month age, information transmitted through social interactions is better processed than information provided by

non-social sources. Furthermore, Vouloumanos et al.'s (2014, p. 877) experiments indicate that, by 6 months, «infants understand that the form of speech, independent of any specific lexical content, can communicate information about an object». It seems there is a rudimentary comprehension of the referential nature of gaze shift, that is facilitated by previous mutual eye contact. Csibra and Volein (2008) suggest that infants can infer that mutual eye contact and the following gaze shift imply a message transmitted by the communicator in front of them, and that such message is the referent. Senju and Csibra (2008) found in 6-month-old infants that such facilitation is evident if the act of following adult's gaze shift is preceded by ostensive signals joint together, as eye contact plus infant-direct speech (IDS), termed also "motherese". Several ostensive signals can be combined. «Ostensive signals *obligatorily indicate* to young infants that communication is directed to them» [my italics] (Parise&Csibra 2013, p. 1). Therefore, once understood that the message is addressed to them, infants are biased to trigger a referential expectation that represents a crucial and preparatory moment to approach the informative content of the message.

Neural areas involved in processing ostensive cues are comparable with adults during ontogenetic development in the first months after birth, but there is not a perfect match under different conditions and modalities of data recording. Therefore, *it is better to adopt a cautious explanation that does not predict a primary form of mindreading involved in the interpretation of communicative intentions by the young recipients*. An alternative, plausible and parsimonious hypothesis can be sustained: through the "fast-track modulator" model proposed by Senju and Johnson (2009), infants early detect communicative signals normally provided through multimodal interactions as mutual eye contact and particular vocal intonations. In virtue of such detecting processes, infants can comprehend in some circumstances that there is a message for them.

This proposal answers the first question posed at the beginning of the present thesis (see Introduction) concerning one aspect of the problematic relation between natural pedagogy and mindreading system. The communicative intentions manifested by the adults/teachers are not interpreted by infants in force of alleged capacities to read others' minds. For triggering the 'pedagogical learning' it is sufficient that infants are well prepared to pick up the full referential potential of ostensive signals. In this way,

they are ready to learn about the world from expertise conspecifics. The ostensive cues given by the adult/teacher/demonstrator induce in naïve learners a special interpretational stance making them ready to acquire novel information that can also violate the principle of efficiency[3] but that are equally deemed rational and relevant to be *learned quickly and permanently*.

3. The cognitive biases of the pedagogical stance

Pedagogical stance shapes inferential processes about epistemic information along two reasoning pathways (or biases): one directed towards the object categorisation, and the other one towards the adults and social contexts. The first inferential pathway is named «assumption of generalizability», whereas the second pathway is termed «assumption of universality» (Gergely et al. 2007, p. 141). Through the former, infants learn not just episodic and local facts «but the generic structure of their cultural worlds» (Tomasello 2016, p. 643). Such bias leads infants to infer that «the pedagogically manifested information about the referent is generalizable to the object kind that the referent belongs to» (Gergely 2007a, p. 179). Therefore, *generalizability* means that the object functions learned in pedagogical context are generalised into a proper kind, in order to be used in other future contexts. In this way, what infants are learning does not only regard the local and single object, but also the categories which objects belong to.

Instead, the *assumption of universality* (termed also *omniscience assumption*, or simply *shared knowledge assumption*) is a bias according to which whatever the child learns (at least) pedagogically is supposed to be known by everyone. Such corollary of omniscience allows that the knowledge that the child acquires is *believed* as public, shared, and universal. «If someone knows something, everyone knows it» (Csibra&Gergely 2006, p. 273). In other words, such assumption triggers the bias whereby what is just learned by infants is alleged by them to be already known by everyone. Quoting Gergely, the universality bias elicits «the implicit expectation by the infant that the manifested information will contain publicly shared universal cultural

[3] I refer here to the teleological reasoning in infancy that Gergely and Csibra described in several essays (I unfold this topic in more details in Chapter 1).

knowledge available to all others (and not only to the demonstrator, who is the communicative source of the information)» (Gergely 2007a, p. 179).

The second question posed at the beginning of my research concerns how *assumption of universality* can occur. In my opinion, the bias of omniscience represents the *bridge that connects in a cooperative way natural pedagogy system and mindreading system*. This implies a partial revision of the original formulation of the theory, because Gergely and Csibra have always denied any links between the two systems. The advocates of the theory of natural pedagogy do not claim that infants develop mindreading skills after natural pedagogy, or that infants (around the first year of age) are unable to ascribe epistemic mental state, emotions, intentions, desires, or dispositional attitudes. Rather, the two Hungarian psychologists affirm an *independence* between pedagogical processing and mentalizing skills to infer and represent mental states of others. I report Gergely's thought: «*The central claim* that pedagogy theory makes *is independent* [from] the question of whether and/or at what point in development young infants are able to infer and represent mental states of others» [my italics] (Gergely 2007a, p. 193). The conclusion prefigured by Gergely is that «one can learn from other minds without learning about them» (Ibidem).

Gergely is apparently justified to sustain this position, because he and Csibra have in mind the simulationist account of mindreading adapted by Andy Meltzoff in his model for social learning called «*Like-me hypothesis*» (see Chapter 3, § 1.3). One of the crucial differences between the theory of natural pedagogy and the *like-me hypothesis* consists in the interpretation given to the degree of kinship between imitation and understanding others' mind. In Meltzoff's perspective, imitation and mindreading are *causally* related. Babies have none folk-psychology skill at birth, but they are equipped with «an imitative brain»; the cultural context and social interactions in which infants are immersed together with psychological agents help the early maturation of *an intentional Self*, that decodes others' actions through the "like-me" process. *The full-blown mindreading would be the product of such social interplay*. The notion "like-me" means that infants see, or rather they recognise others like themselves. «*Human* acts are *especially relevant* to infants because they *look like* the infant *feels* himself to be and because they are events that infants can *intend*» [my italics] (Meltzoff 2005, p. 74). Therefore, according to Meltzoff, seeing a human act means to infants recognising an

event previously felt in the body. The "like-me" hypothesis suggests that the infants map the other from the Self.

On the contrary, the pedagogical learning deals with an object-centred perspective that ignores (or fails) the teacher's mental state. Thus, it would be exactly the opposite of a person-centred perspective predicted by the simulationist mindreading account. Indeed, such assumption implies that an individual behaviour manifested pedagogically as a personal dispositional property is learned by infants and children as an epistemic status (or a knowledge content) that goes beyond the episodic situation in which it has been manifested to be relevantly ascribed to other members of one's social group.

According to Gergely, the comprehension of agent's behaviour is primarily tied not to the agent herself, intended as the target of the observer's intentional projections, rather to the structure of intentional action performed and its relative context with environmental constraints. From an epistemological point of view, as I have emphasised in Chapter 1, this entails the recruitment of top-down processing for action interpretation, rather than bottom-up processes as suggested by the simulationist account *à la* Meltzoff.

4. The supposed modalities of cooperation between natural pedagogy and mindreading

I deem that the universality assumption is crucial for the efficacy of transmission and above all for the maintenance of cultural knowledge under the form of conventions and common ground beliefs between generations. In virtue of natural pedagogy «human children do not just culturally learn useful instrumental activities and information, *they conform to the normative expectations of the cultural group* and even contribute themselves to the creation of such normative expectations» (Tomasello 2016, p. 643).

According to the assumption of universality, the knowledge content assimilated pedagogically is, thus, ascribed to others. I suggest that this attribution is permitted in virtue of a primary form of mindreading that is extremely flexible and compatible with the cognitive architecture sketched by Belief file's theoretical construal (Chapter 5, § 3). Proposed by Kovács (2016), such belief attribution 'machinery' enables to track and update the own representations about others' beliefs. *Belief file* is defined as the «basic

representational structure» that operates from the early infancy enabling «implicit ToM processes to store information about other agents' beliefs in a format supporting efficient encoding and updating» (Kovács 2016, p. 510). Its structure is flexibly articulated in three horns (the agent as belief-holder, the belief-content, and the referent). Two are the variables that can be separately updated giving the flexible profile to whole structure: 1) the belief-holder and 2) the belief content. However, only the former is necessary to permit the beginning of propositional attitudes attribution. In other words, an agent as target of belief attribution must be present. Subsequently, the belief holder can be replaced by «an "agent–placeholder"» to which infants can attribute the same belief content (Kampis et al. 2013, p. 238). In this way, the content of a message is not constrained by the contingent circumstances and can be attributed freely to others.

By decomposing belief attribution processing, we may distinguish the starting phase of opening the belief file from the computations of its contents, and the bond between belief representations and the corresponding agents[4]. This permits to attribute the achieved and stored information to other subjects in further circumstances. Such precocious flexible form of epistemic contents ascription may constitute the cognitive basis for the omniscience bias occurred in pedagogical relationship.

The amount of evidence collected by several findings presented in Chapter 4 and 5 indicate two important elements:

1) the relevant recruitment of false belief reasoning for learning something about the world through others' perspective. Given the close connection between *world-representations* and *others' false beliefs-representation,* and the influence employed by the latter on the former, infants may gain information about the world from (false) belief attribution practise in specific circumstances. Precocious infants' use of false belief attribution reveals a lot about the very nature of belief "handled" by infant's mind. Therefore, given particular constraints, *false belief computation serves to achieve*

[4] In Chapter 5 (§ 3.2 and followers), I emphasised the (not only terminological) link between Belief-file and Recanati's Mental file account, that represents in this context an useful theoretical tool for understanding the starting phase of opening the belief file insofar mental file belongs to the system of mental representations and refers to them as *modes of presentation* of the object in Fregeanian sense.

information about the world, and by this acquisition to *predict* others' behaviour. The flexibility of early mental representation attribution reveals its powerful social adaptive advantage in the lack of binding between belief-holder and the conveyed epistemic content, which can be reused by infants as «a shared knowledge […] applicable to other agents as well» (Kampis et al. 2013, p. 238). Furthermore, «the lack of binding of mental states to agents […] could serve an important role in promoting joint action and cooperation» (Kampis et al. 2013, p. 238).

2) A conceptual interpretation of others' behaviour based on propositional attitudes.

5. Propositional nature of infant beliefs

Representations are inferentially linked each other and have a *propositional format* like regular beliefs. It is normally accepted the definition of belief as states that *represent* how things stand in the world (Schwitzgebel 2015). In force of their propositional form, beliefs involve conceptual contents, or in other terms, the explicit form of the propositional stance has implicit in it the conceptual content of the proposition articulated. Therefore, within the notion of *belief*, conceptual, propositional and representational dimensions are linked each other in depth. On this premise, infants build their expectations as *conceptual creatures*. Therefore, grasping propositional contents does not imply the necessary involvement of a linguistic dimension. Infants do not wait to become speakers to be believers (see Chapter 5, § 1.3).

Entertaining regular beliefs entails the notion of *aspectuality*, namely the capacity to entertain *simultaneously* a dual identity representation relative to the same target object. Belief attribution machinery shows very early a great flexibility in several tasks also those encompassing dual identity object recognition, as we have seen in Chapter 5 (§ 4.4) by presenting the recent findings of Buttelmann et al. (2015) and Kampis (2017). According to their results, infants distinguish the double identity of the objects, and such accomplishment enables them to use person's belief about the object identity in order to infer his/her goal and to help his/her accordingly. Beyond aspectuality competence infants show remarkable flexibility not only for attributing false beliefs about location, but also for attributing other reality-incongruent epistemic states like pretence (e.g., Onishi et al. 2007), false perceptions (Song&Baillargeon 2008), false

information about non-obvious properties (Scott et al. 2010), false information about identity (Scott&Baillargeon 2009).

On the basis of the findings presented in the previous chapters, I think it is reasonable to abandon the hypothesis of a *representational discontinuity* throughout the development as predicted by the Apperly and Butterfill's two-system model of mindreading (Chapter 5, § 2). Mindreading may be seen as an output of a complex cooperation among several human cognitive mechanisms underlying pragmatic competences, and executive functions, working memory system, and eventually the competitive cooperation with natural pedagogy too.

The salience information obtained from a false belief condition can represent in this way a precious information to use in further social interactions. I deem that such general tendency to attribute universally epistemic contents can be accommodated for more sophisticated social learning context, as pedagogical contexts actually are. *In this way, the natural pedagogy system makes use of this innate capacity that is already in place between 7 and 10 months of age, and employed in other kind of social interactions.*

I opened the present thesis with two questions (Introduction, p. 11):

i) Does natural pedagogy theory claim that a communicative intention is not part of theory of mind?
ii) Does natural pedagogy theory claim that the ascription of informative contents to others can be not part of mindreading system?

For the question i), I suggest that natural pedagogy *does not 'use' precocious TOM skills* to interpret communicative intentions, which are decoded by other deep processes. For the question ii), I suggest that *natural pedagogy utilises belief attribution competences which are employed by infants in a variety of context for approaching and navigating the social world.*

6. Bounder of knowledge domain via natural pedagogy system

All these findings induce to think that infants see and look for other people around them not only as a source of nutritive and emotional care, but also as a source of information about the surrounding environment, and as guides who allow them to learn about natural and artefact objects.

I have assumed, in this way, a sort of *innate epistemic attitude* that projects infants towards the world like spontaneous searchers, or hunters of information and cultural knowledge potentially obtainable by any kind of communicative sources able to catch first and shift then their attention towards an external referent. However, in the last chapter, I have fixed the bounders of the knowledge domain available via natural pedagogy system. Among the wide pragmatic territory of knowledge, the inner emotional states are not included. Against Gergely, Unoka and other influential colleagues, I suggest that the self-awareness of feeling emotions - or as constructivist theorists *à la* Barrett (2006a; 2006b)[5] say - the "*conceptualisation of emotions*" is not the result of a pedagogical activity committed with social biofeedback apparatus (conceived originally by Gergely and Watson (1996; 1999)[6]). The affective dyadic relation establishing between adult caregivers and infants is not representable as, or reducible to, an intentional teaching act to which a learning practice can correspond with the subsequent cognitive biases that I have illustrated so far.

7. Values and limits of natural pedagogy

Teaching is always a form of (intentional) communication. If natural pedagogy can be defined without any doubt as a "direct active teaching" (DAT) akin to other teaching modalities observed among other non-human animals, at the same time, natural pedagogy is human-specific among social learning practices because the teacher does not merely play the role of attractor of pupil's attention towards certain actions, or objects. Rather, through ostensive cues, the teacher «manifests to the pupil that she is the intended addressee of the demonstration» (Tanone&Csibra 2015, p. 50).

[5] See Chapter 7 for more details and references.

[6] I have illustrated the theory of social biofeedback in Chapter 7.

Acknowledging the existence of DAT in non-human animal kingdom and the use of referential signals (e.g., alarm calls, food calls, bee dance, etc.), human ostensive communication is unique because it allows acquiring generic knowledge contents that are opaque material and social kinds, artifacts and conventions, not functionally transparent, and without any evident fitness value. The evolution of such cognitive adaptations allowed pupils to interpret the taught information as being applicable *beyond its episodic use and contextual achievement*, whereas the referential signals used in non-human DAT are «restricted to episodic facts in the "here and now" and cannot express content that is generalizable to other situations, other locations or other individuals» (Csibra&Gergely 2011, p.1150).

The theory of natural pedagogy has been attacked, as I explain in Chapter 6, for its supposed identification with ostensive communication. I clarify that if ostensive communication is necessary for establishing pedagogical relations, the opposite is not the case. In natural pedagogy theory the *credibility* is given through the communicative modalities of transmission, i.e., *how* the testimony is communicated, namely through ostensive communication. Natural pedagogy predicts that the ostensive nature of communicative transmission tend to drop any need for evaluation and allows fast and efficient learning. However, as a series of evidence presented in Chapter 6 shows, children's trust in attester is not granted. Rather, it is a result deeply affected by a dynamic and conflicting balance between the inclination to trust blindly induced by the teacher's communicative approach and the increasing competence of reasoning together with abilities to read others' intention and mental states. Pupils are able to select their informative sources from a very early stage. This capacity increases significantly in social contexts during the development and leads children to inhibits the potential factors which trigger pedagogical stance.

Natural pedagogy provides a model of learning by trust. Children learn more efficiently by trusting and, subsequently, by believing the adult through the transmission of testimonies. The knowledge achieved in this way forms a system of convictions that constitute common ground of beliefs. Definitively, *we can define natural pedagogy model as a developmental temporary form of cross-generational transmission of testimonies*, where the notion of 'testimony' represents – psychologically speaking - «a

source of basic beliefs» (Audi 2015, p. 230). The concept of common ground is dramatically important for navigating properly in the social environment, understanding others' behaviour, and sustaining adequately conversations.

In conclusion, I hope to be clear the crucial role I suppose the natural pedagogy system plays in the construction of a stable network of beliefs relatively to some domains. I have tried to border the contextual application of natural pedagogy that represents only one of social learning strategies employed by "teachers" and pupils. Natural pedagogy system, based on child-directed interactions and in cooperation with the early mindreading system, can represent one of the most efficient adaptive strategies to create firmly that deep «nest of propositions» which - Wittgenstein (1949/1969, § 102; § 225) wrote in *On Certainty* - build cultural shared beliefs structure to be relied on and followed. We may claim that through natural pedagogy children can learn a host of things belonging to cultural knowledge domain, and they learn to act according to such beliefs.

Acknowledgements

I have to thank many people. Some of them have become friends. I would like to start by some teachers of Consortium FINO, and in particular Carlo Penco (University of Genoa), Alberto Voltolini (University of Turin), Carola Barbero (University of Turin), Alfredo Paternoster (University of Turin), prof. Pasqaule Frascolla (University of Basilicata), Cristina Amoretti (University of Genoa), Andrea Iacona (University of Turin), Massimiliano Vignolo (University of Genoa), Diego Marconi (University of Turin), Marcello Frixione (University of Genoa). Their suggestions have been always relevant during these four years.

Sincere gratitude to my colleagues, Marco Viola, Marta Benenti, Enirco Grosso, Enrico Cipriani, Davide Serpico. Thanks also to Jlenia Quartarone and Richard Moore for their suggestions.

During the first period of PhD I could attend stimulating lessons at IUSS, in Pavia. I thank a lot prof. Michele Di Francesco, Andrea Moro and Valentina Bambini.

I sincerely thank the amazing and dynamic group that composes CEU Cognitive Department in Budapest. And in particular my special thanks go to Dan Sperber, who met for the first time in Pavia at IUSS, Dora Kampis, Agnes M. Kovács, Nicolò Cesana-Arlotti, Pierre Jacob, and of course György Gergely and Gergely Csibra who allowed me to attend their fascinating lessons. Without their hard and fruitful work, this dissertation would not exist. I wish you all the best.

I have been very lucky to have evaluators like Carlo Lai (University of Rome "La Sapienza") and Pietro Perconti (University of Messina), whose suggestions have been fundamental for the actual and future works.

A very special thank goes to my tutor, Cristina Meini (University of East Piedmont) and to Massimo Marraffa (University of "Roma Tre"). The study of their papers and books has been always illuminating. They always supported my research and indicated intriguing paths of philosophical analysis.

Finally, my family, for the patience and encouragement. And thanks to you, my friend, you have been always present together with our two little dreams always in our mind.

References

Adams F., Aizawa K. (2017) Causal Theories of Mental Content, In: Zalta E.N. (ed.) *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/sum2017/entries/content-causal/>.

Adler J.E. (1994) Testimony, Trust, Knowing. *Journal of Philosophy*, 91, 264-275.

Amoretti M.C. (2012) Sul carattere sociale di pensiero e linguaggio. *RIFL. Rivista Italiana di Filosofia del Linguaggio*, Numero Speciale SFL, 4–16.

Amoretti M.C. (2013) Concepts within the model of triangulation. *ProtoSociology. An International Journal of Interdisciplinary Research*, 30, 49-63.

Andrews K. (2016) Pluralistic Folk Psychology in humans and other apes. In Kiverstein J. (ed.) *The Routledge Handbook of Philosophy of the Social Mind* (pp. 117-138). London, Routledge.

Apperly, I.A. (2011) *Mindreading*. New York, Psychology Press.

Apperly, I.A., Robinson E.J. (1998) Children's mental representation of referential relations. *Cognition*, 67, 287–309.

Apperly, I.A., Robinson E.J. (2003) When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year-old children's reasoning about beliefs and belief reports. *Journal of Experimental Child Psychology*, 85(4), 297-311.

Apperly I.A., Samson D., Chiavarino C., Humphreys G.W. (2004) Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, 16, 1773–1784.

Apperly I.A., Butterfill S.A. (2009) Do humans have two systems to track beliefs and belief-like states? *American Psychological Association*, 116(4), 953-970.

Audi R. (2015) *Rational beliefs. Structure, grounds, and intellectual virtue*. Oxford, Oxford UP.

Baars, B.J. (1988) *A cognitive theory of consciousness*. Cambridge University Press.

Baars B.J., Ramsoy T., Laureys S. (2003) Brain, consciousness, and the observing self. *Trends in Neurosciences*, 26, 671–75.

Baillargeon R. (2008) Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, 3, 2-13.

Baillargeon R, Scott R.M., He Z. (2010) False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110–118.

Baillargeon R., Stavans M., Wu D., Gertner Y., Setoh P., Kittredge A.K., Bernard A. (2012) Object Individuation and Physical Reasoning in Infancy: An Integrative Account. *Language Learning and Development*, 8, 4–46.

Baillargeon R., Scott R.M., He Z., Sloane S., Setoh P., Jin K., Wu D., Bian L. (2015) Psychological and sociomoral reasoning in infancy. In Mikulincer M., Shaver P.R., Borgida E.,

Bargh J.A. (eds.) *APA Handbook of Personality and Social Psychology*, vol. 1. *Attitudes and Social Cognition* (pp. 79–150). Washington, DC, American Psychological Association.

Baillargeon R., Scott R.M., Bian L. (2016) Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 159-186.

Baldwin D. (1991) Infants' contribution to the achievement of joint reference. *Child Development*, 62, 875–890.

Baldwin D. (1993) Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, 64, 711–728.

Bar-On D. (2013) Origins of communication: Must we "Go Gricean?". *Mind & Language*, 28, 342–75.

Bar-On D., Priselac M. (2011) Triangulation and the Beasts. In Amoretti M.C., Preyer G. (eds.) *Triangulation. From an Epistemological Point of View.* (pp. 121–152). Frankfurt, Ontos Verlag.

Bardi L., Desmet C., Nijhof A., Wiersema J.R., Brass M. (2017) Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience*, 12(3), 391-400.

Baron-Cohen S., Leslie A. M., Frith U. (1985) Does the autistic child have a "theory of mind?" *Cognition*, 21, 37–46.

Baron-Cohen S. (1994) How to build a baby that can read minds: cognitive mechanisms in mindreading. *Current Psychology of Cognition*, 13, 1–40.

Baron-Cohen S. (2005) The Empathizing System: a revision of the 1994 model of the Mindreading System. In: Ellis B., Bjorklund D. (eds) *Origins of the Social Mind* (pp. 468-492) Guilford Publications Inc.

Barrett H.C., Broesch T., Scott R.M., He Z., Baillargeon R., Wu D., Bolz M., Joseph Henrich, Setoh P., Wang J. Laurence S. (2013) Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B*, 280, 20122654, 1-6.

Barrett L.F. (2006a) Are emotions natural kinds? *Perspectives on Psychological Science*, 1(28), 28-58.

Barrett L.F. (2006b) Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, 10(1), 20-46.

Barrett L.F., Wilson-Mendenhall C.D., Barsalou L.W. (2015) The Conceptual Act Theory: A road map. In: Barrett L.F., Russell J.A. (eds.), *The Psychological Construction of Emotion* (pp. 83-110). New York, Guilford Press.

Bas J., Sebastián-Gallés N. (2016) Influence of social dominance on learning. In *BCCCD 2016. Budapest CEU Conference on Cognitive Development*, Budapest 07-09 Janaury 2016, p. 62

Batki A., Baron-Cohen S., Wheelwright S., Connellan J., Ahluwalia J. (2000) Is there an innate gaze module? Evidence from human neonates. *Infant Behavior & Development*, 23, 223–229.

Becchio C., Koulab A., Ansuinia C., Cavallo A. (2018) Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of Life Review*, 24, 67-80.

Bernecker S. (2013) Triangular Externalism. In: Lepore E., Ludwig K.(eds.), *A Companion to Donald Davidson* (pp. 443-455), Hoboken, NJ, Wiley Blackwell.

Bertin E., Striano T. (2006) The still-face response in newborn, 1.5-, and 3-month-old infants. *Infant Behavior & Development*, 29, 294-297.

Biervoje A., Dricot L., Ivanoiu A., Samson D. (2016) Impaired spontaneous belief inference following acquired damage to the left posterior temporoparietal junction. *Social Cognitive and Affective Neuroscience*, 1-8.

Bigelow A. E. (2009) Self knowledge. In Benson J.B., Haith M.M. (Eds.), *Social and emotional development in early childhood* (pp. 353-364). Oxford, UK: Elsevier.

Bigelow A. E., De Coste C. (2003) Sensitivity to social contingency from mothers and strangers in 2-, 4-and 6-month-old infants. *Infancy*, 4, 111-140.

Birsch S.A., Vauthier S.A., Bloom P. (2008) Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, 107(3):1018-34.

Birsch S.A.J., Li V., Haddock T., Ghrear S.E., Brosseau-Liard P., Baimel A., Whyte M. (2017) Perspectives on perspective taking: How children think about the minds of others. *Advances in Child Development and Behavior*, 52,185-226.

Block N. (1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227–247.

Bloom L., Rocissano L., Hood L. (1976) Adult–child discourse: Developmental interaction between information processing and linguistic knowledge. *Cognitive Psychology*, 8, 521–552.

Bloom P. (2004) *Descartes'baby. How the science of child development explains what makes us human*. New York, Basic Books.

Bloom P. (2010) *How pleasure works. The new science of why we like what we like*. Norton New York-London Norton & Company.

Bloom P., German T. P. (2000) Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77(1), 25–31.

Boesch C., Tomasello M. (1998) Chimapazee and human cultures. *Current Anthropology*, 39, 591-614.

Bonawitz E., Shafto P., Gweon H., Goodman N., Spelke E., Schulz L. (2011) The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120(3), 322-330.

Booth A.E., Waxman S.R. (2002) Word learning is 'smart': evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, 84, B11-B22.

Bosbach S., Cole J., Prinz W., Knoblich G. (2005) Inferring another's expectation from action: The role of peripheral sensation. *Nature Neuroscience*, 8(10), 1295–1297.

Bosco F.M., Bucciarelli M., Bara B.G. (2004) The fundamental context categories in understanding communicative intention. *Journal of Pragmatics*, 36, 467–488.

Bosco F.M., Friedman O., Leslie A.M. (2006) Recognition of pretend and real actions in play by 1- and 2-Year-Olds: Early success and why they fail. *Cognitive Development*, 21(1), 3-10.

Brand R., Baldwin D., Ashburn L. (2002) Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, 5(1), 72-83.

Brand R., Shellcross W.L. (2007) Infants prefer motionese to adult-direct-action. *Developmental Science*, 11, 853-861.

Brandom R. (2000) *Articulating reasons. An introduction to inferentialism*. Cambridge, Harvard UP.

Brauer J., Kaminski J. R., Riedel J., Call J., Tomasello M. (2006) Making inferences about the location of hidden food: Social dog, causal ape. *Journal of Comparative Psychology*, 120 (1), 38 – 47.

Breheny R. (2006) Communication and folk psychology. *Mind & Language*, 21(1), 74-107.

Brink I. (2004) Joint Attention, triangulation and radical interpretation: A problem and its solution. *Dialectica*, 58, 179–206. (2005)

Brooks R., Meltzoff A.N. (2005) The development of gaze following and its relation to language. *Developmental Science*, 8, 535-543.

Bruner J. (1977) Early social interaction and language acquisition. In H. Schaffer (Ed.), *Studies in mother–infant interaction* (pp. 271–289). London, Academic Press.

Buchsbaum D., Gopnik A., Griffiths T.L., Shafto P. (2011) Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, 120, 331-340.

Bugnyar T. (2011), Knower-guesser differentiation in ravens: others' viewpoint matters. *Proceeding of the Royal Society B: Biological Sciences*, 278(1705), 634-640.

Bugnyar T., Stowe M., Heinrich B. (2004) Ravens, Corvus corax, follow gaze direction of humans around obstacles. *Proceedings of the Royal Society of London, Series B*, 271, 1331-1336.

Buresh J.S., Woodward A.L. (2007) Infants track action goals within and across agents. *Cognition*, 104(2), 287–314.

Burkart J.M., Hrdy S.B., Van Schaik (2009) Cooperative Breeding and Human Cognitive Evolution. *Evolutionary Anthropology*, 18, 175–186.

Burnside K., Ruel A., Azar A., Poulin-Dubois D. (2017) Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development*, available on-line https://doi.org/10.1016/j.cogdev.2017.08.006

Butler L.P., Markman E.M. (2012) Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, 83(4), 1416–1428.

Butler L.P., Markman E.M. (2014) Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*, 130(1), 116–127.

Butler L.P., Schmidt M.F.H., Brügel J., Tomasello M. (2015) Young children use pedagogical cues to modulate the strength of normative inferences. *British Journal of Developmental Psychology*, 33(4), 476-488.

Buttelmann D., Carpenter M., Call J., Tomasello M. (2008) Rational tool use and tool choice in human infants and great apes. *Child Development*, 79(3), 609–626.

Buttelmann D., Carpenter M., Tomasello M. (2009) Eighteenmonth-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112, 337–342.

Buttelmann D., Zmyj N. (2012) Evaluating the empirical evidence for two-stage model of infant imitation. A commentary on Paulus, Hunnius, Vissers, and Bekkering. *Frontiers in Psychology*, 3, 512.

Buttelmann D., Zmyj N., Daum M., Carpenter M. (2013) Selective Imitation of In-Group Over Out-Group Members in 14-Month-Old Infants. *Child Development*, 84(2), 422-428.

Buttelmann D., Over H., Carpenter M., Tomasello M. (2014) Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology*, 119, 120–126.

Buttelmann F., Suhrke J., Buttelmann D. (2015) What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, 131, 94–103.

Buttelmann D., Buttelmann F., Carpenter M., Call J., Tomasello M. (2017) Great apes distinguish true from false beliefs in an interactive helping task. *PlosOne*, 12(4), 1-8.

Butterfill S.A., Apperly I.A. (2013) How to construct a Minimal Theory of Mind. *Mind & Language*, 28(5), 606-637.

Byrne A. (2005) Introspection. *Philosophical Topics*, 33, 79-104.

Call J., Tomasello M. (2005/2010) What Chimpanzees know about seeing, revisited: an explanation of the third kind. In: Elian N., Hoerl C., McCormack T., Roessler J. (eds.) *Joint Attention: Communication and other minds*. Issues in Philosophy and Psychology Oxford, Clarendon press. Published on-line in 2010. DOI:10.1093/acprof:oso/9780199245635.003.0003

Call J., Tomasello M. (2008) Does the chimpanzee have a theory of mind? Thirty years later. *Trends in Cognitive Science*, 12, 187 – 192.

Call J., Tomasello M. (2009) A Nonverbal False Belief Task: The performance of children and great apes. *Child Development*, 70, 381–395.

Callaghan T., Rochat P., Lillard A., Claux M.L., Odden H., Itakura S., Tapanya S., Singh S. (2005) Synchrony in the onset of mental-state reasoning: evidence from five cultures. *Psychological Science*, 16(5), 378-84.

Cannon E.N., Woodward A.L. (2012) Infants generate goal-based action predictions. *Developmental Science*, 15(2), 292-298.

Carey S. (2009) *The origins of concepts*, Oxford, Oxford UP.

Caro T.M., Hauser M.D. (1992) Is there teaching in nonhuman animals? *Quarterly Review of Biology*, 67(2), 151–174.

Caron A.J. (2009) Comprehension of the representational mind in infancy. *Developmental Review*, 29, 69–95.

Carpenter M., Nagell K., Tomasello M. (1998) Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63, 1–174.

Carrington S.J., Bailey A.J. (2009) Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, 30(8), 2313–2335.

Carruthers P. (2009) How we know our own minds: The relationship between mindreading and metacognition. *Behavioral And Brain Sciences*, 32, 121–182.

Carruthers P. (2011) *Opacity of mind. An integrative theory of Self-Knowledge*. Oxford, Oxford UP.

Carruthres P. (2013a) Mindreading in infancy. *Mind & Language*, 27(2), 141-172.

Carruthers P. (2013b) Mindreading the Self. In: S. Baron-Cohen, M. Lombardo, H. Tager-Flusberg, *Understanding Other Minds: Perspectives from developmental social neuroscience*, Oxford, Oxford Scholarship Online. My quotations refer to the online version available. DOI:10.1093/acprof:oso/9780199692972.003.0026.

Carruthers P. (2015) *The centered mind: what the science of working memory shows us about the nature of human thought*, Oxford UP, Oxford.

Carruthers P. (2016) Two systems for mindreading? *Review of Philosophy and Psychology*, 7(1), 141–162.

Carruthers P. (2017) Mindreading in adults: evaluating two systems views. *Synthese*, 194, 673–688.

Caruana F. (2017) Dal costruzionismo alla teoria motoria delle emozioni. *Sistemi Intelligenti*, 29(1), 145-167.

Caruana F., Viola M. (2018) *Come funzionano le emozioni. Da Darwin alle neuroscienze*. Bologna, Il Mulino.

Carver L.J., Dawson G., Panagiotides H., (2003) Age-related differences in neural correlates od face recognition during the toddler and preschool years. *Developmental Psychobiology*, 42, 148-159.

Catmur C. (2013) Sensorimotor learning and the ontogeny of the mirror neuron system. *Neuroscience letters*, 540, 21-27.

Cecchini M., Baroni E., Di Vito C., Piccolo F., Lai C. (2011) Newborn preference for a new face vs. a previously seen communicative or motionless face. *Infant Behavior and Development*, 34, 424-433.

Cecchini M., Lai C., Langher V. (2008) Communication and crying in newborns. *Infant Behavior & Development*, 30, 655–665.

Cesana-Arlotti N., Martín A., Téglás E., Vorobyova L., Cetnarski R., Bonatti L. (2018) Precursors of logical reasoning in preverbal human infants. *Science*, 359(6381), 1263–126.

Chittka L., Rossiter S.J., Skorupski P., Fernando C. (2012) What is comparable in comparative cognition? *Philosophical Transactions of the Royal Society B*, 367(1603), 2677-2685.

Christensen W., Michael J. (2016) From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, 40, 48-64.

Cimpian A., Markman E.M. (2009) Information learned from generic language becomes central to children's biological concepts: evidence from their open-ended explanations. *Cognition*, 113(1), 14-23.

Cimpian A., Cadena C. (2010) Why are dunkels sticky? Preschoolers infer functionality and intentional creation for artifact properties learned from generic language. *Cognition*, 117(1), 62-68.

Cimpian A., Scott R.M. (2012) Children expect generic knowledge to be widely shared *Cognition*, 123(3), 419-433.

Clark A. (1998) Being There: Putting Brain, Body, and World Together Again. MIT Press

Clark A. (2013) Whatever next? Predictive brain, situated agents, and the future of cognitive science. *Behavioral Brain Science*, 36(3), 181-204.

Clark H.H., Marshall C.R. (1992) Definite reference and mutual knowledge. In H.H. Clark (ed.) *Arenas of Language Use* (pp. 9-39). London, University Chicago Press.

Clément F., Dukes D. (2017) Social appraisal and social referencing: Two components of affective social learning. *Emotion Review*, 9(3), 253-261.

Connellan J., Baron-Cohen S., Wheelwright S., Batki A., Ahluwalia J. (2001) Sex differences in human neonatal social perception. *Infant Behavior and Development*, 23, 113-118.

Connors M.H., Halligan P.W. (2015) A cognitive account of belief: a tentative road map. *Frontiers in psychology*. DOI: 10.3389/fpsyg.2014.01588.

Cooper R.P., Aslin R.N. (1990) Preference for infant-directed speech in the first month after birth. *Child Development*, 61, 1584–1595.

Corriveau, K.H., Harris P.L., Meins E., Fernyhough C., Arnott B., Elliott L., Liddle B., Hearn A., Vittorini L., de Rosnay M. (2009) Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, 80, 750-761.

Crivello C., Poulin-Dubois D. (2017) Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development*, published on line. DOI: 10.1016/j.cogdev.2017.10.003

Csibra G. (2003) Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358, 447-458.

Csibra G. (2007a) Action mirroring and action understanding: An alternative account. In: P. Haggard, Y. Rosetti, M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435-459). Oxford, Oxford University Press.

Csibra G. (2007b) Teaching in the wild. *Trends in Cognitive Sciences*, 11(3), 95-96.

Csibra G. (2010) Recognizing communicative intentions in infancy. *Mind & Language*, 25(2), 141–168.

Csibra G., Gergely G. (1998) The teleological origins of mentalistic action explanations: a developmental hypothesis. *Developmental Science*, 1, 255–259.

Csibra G., Biro S., Koos O., Gergely G. (2003) One-year-old infants use teleological representations of actions productively. *Cognitive Science*, 27(1), 111–133.

Csibra G., Gergely G. (2006) Social learning and social cognition: The case for pedagogy. In Y. Munakata, M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance*, XXI (pp. 249-274). Oxford, Oxford University Press.

Csibra G., Southgate V. (2006) Evidence for infants understanding of false beliefs should not be dismissed. *Trends in Cognitive Sciences*, 10, 4-5

Csibra G., Gergely G. (2007) 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, 124, 60-78.

Csibra G., Volein A. (2008) Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology* (2008), 26, 1–11.

Csibra G., Gergely G. (2009) Natural pedagogy. *Trends in Cognitive Science*, 13(4), 148-153.

Csibra G., Gergely G. (2011) Natural pedagogy as evolutionary adaptation. *Philosophical Transaction of the Royal Society B*, 346, 1149-1157.

Csibra G., Shamsudheen R. (2015) Nonverbal generics: Human infants interpret objects as symbols of object kinds. *Annual Review of Psychology*, 66, 689-710.

Davidson D. (1980) *Essays on Actions and Events*, Oxford, Oxford University Press.

Davidson D. (1991), Three Varieties of Knowledge. In: ID. (2001), *Subjective, Intersubjective, Objective* (pp. 205-220). Oxford, Oxford UP.

Davidson D. (1992) The second person. In: ID. (2001) *Subjective, Intersubjective, Objective* (pp. 107–122). Oxford, Oxford UP.

Davidson D. (1997a) The emergence of thought. In: ID. (2001) *Subjective, Intersubjective, Objective* (pp. 123–134). Oxford, Oxford UP.

Davidson D. (1997b) Seeing through Language. In: ID. (2005) *Truth, Language, and History* (127–142). Oxford, Oxford UP.

Davidson D. (2001) Externalisms. In: Kotatko P., Pagin P., Segal G. (eds.), *Interpreting Davidson* (pp. 1-16). Stanford, CSLI.

Davies, M. (1989) Tacit knowledge and subdoxastic states. In: George A (ed.), *Reflections on Chomsky* (pp. 131–152). Oxford, Blackwell.

De Bruin L.C., Newen A. (2012) An association account of false belief understanding. *Cognition*, 123, 240–259.

De Caro M. (2011) The short happy life of the swampman: Interpretation and social externalism in Davidson. In: Amoretti M.C., Preyer G. (eds.) *Triangulation. From an Epistemological Point of View* (pp. 179-196). Frankfurt, Ontos Verlag.

de Villiers P.A. (2005) The role of language in theory of mind development: what deaf children tell us. In: Astington J.W., Baird J., (eds), *Why language matters for theory of mind* (pp. 266–297). Oxford University Press. New edition (2012) Oxford Scholarship online. DOI: 10.1093/acprof:oso/9780195159912.003.0013.

de Waal F. B. M., Ferrari P. F. 2010. Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Science*, 14, 201–207.

Deligianni F., Senju A., Gergely G., Csibra G. (2011) Automated gaze-contingent objects elicit orientation following in 8-months-old infants. *Developmental Psychology*, 47(6), 1499-1503.

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

DeVries W. (2016) Wilfrid Sellars. In: Zalta E.N. (ed.) *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2016/entries/sellars.

Dickerson K., Gerhardstein P., Zack E., Barr R. (2013) Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology*, 55(7), 719-732.

Di Francesco M., Marraffa M., Paternoster A. (2016) *The Self and its defenses. From Psychodynamics to Cognitive Science*, London, Palgrave McMillan.

Domaneschi F., Penco C. (2017) Presupposizioni. *Aphex*, 15, Available online http://www.aphex.it/index.php?Temi=557D030122027403210403027773 27.

Dominey P.F., Prescott T.J., Bohg J., Engel A.K., Gallagher S., Heed T., Hoffmann M., Knoblich G., Prinz W., Schwartz A. (2016) Implication of action-oriented paradigm shifts in cognitive sciences. In: Engel A.K., Friston K.J., Kragic D. (eds.) *The Pragmatic Turn: Toward action-oriented views in cognitive science*. *Strüngmann Forum Reports*, 18 (pp. 333-356). Cambridge MA, MIT Press.

Egyed K., Király I., Gergely G. (2004) *Object-centered versus agent-centered Interpretations of referential attitude expressions in 14-month-olds*. Poster presented at the 14th Biennial International Conference on Infant Studies, May 2004, Chicago, IL.

Egyed K., Király I., Krekó K., Kupán K., Gergely G. (2007) *Understanding object referential attitude expressions in 18-month-olds: The interpretation switching function of ostensive communicative cues*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Boston, MA, March 29–April 1, 2007.

Egyed K., Király I., Gergely G. (2013) Communicating shared knowledge in infancy. *Psychological Science*, 24(7), 1348–1353.

Ekman P. (1992) Are there basic emotions? *Psychological Review*, 99(3), 550-553.

Elsabaggah M., Fernandes J., Webb S.M., Dawson G., Charman T., Johnson M.H., British Autism Study of Sibling Team (2013) Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, 74(3), 189-94.

Emery N. J., Clayton N. S. (2001) Effects of experience and social context on prospective caching strategies by scrub jays. *Nature*, 414, 443 – 446.

Engel A.K, Maye A., Kurthen M., Konig P. (2013) Where's the action? The pragmatic turn in cognitive science, *Trends in Cognitive Sciences*, 17(5), 202-9. doi: 10.1016/j.tics.2013.03.006.

Enrici I. Adenzato M., Cappa S., Bara G., Tettamanti M. (2011) Intention processing in communication: A common brain network for language and gestures. *Journal of Cognitive Neuroscience*, 23(9), 2415-2431.

Farroni T., Johnson M.H., Brockbanck M., Simion F. (2000) Infant's use of gaze direction to cue attention: the importance of perceived motion. *Visual Cognition*, 7, 705-718.

Farroni T., Csibra G., Simion F., Johnson M.H. (2002) Eye contact detection in humans from birth. *PNAS*, 99(14), 9602-9606.

Farroni T., Mansfield E., Lai C., Johnson M.H. (2003) Perceiving and acting on the eyes: Tests and evolutionary hypothesis. *International Journal of Experimental Child Psychology*, 85, 199–212.

Farroni T., Johnson M.H., Csibra G. (2004) Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, 16, 1320–26.

Farroni T., Menon E., Johnson M. H. (2006) Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology*, 95, 298–308.

Farroni T., Massaccesi S., Menon E., Johnson M.H. (2007) Direct gaze modulates face recognition in young infants. *Cognition*, 102, 396–404.

Fenici M., Garofoli D. (2017) The biocultural emergence of mindreading: Integrating cognitive archaeology and human development. *Journal of Cultural Cognitive Science*, in press.

Fizke E., Butterfill S., van de Loo L., Reindl E. Rakoczy H. (2017) Are there signature limits in early theory of mind? *Journal of experimental Child Psychology*, 162, 209-224.

Fodor J.A. (1975) *The Language of Thought*. Cambridge, Harvard University Press.

Fodor J.A. (1983) *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MIT Press.

Fonagy P., Gergely G., Jurist E., Target M. (2002) *Affect-regulation, mentalization, and the development of the self*, New York, Other Press.

Fonagy P., Gergely G., Target M. (2007) The parent-infant dyad and the construction of the subjective Self. *Journal of Child Psychology and Psychiatry*, 47(3-4), 288-328.

Fonagy P., Luyten P., Allison E. (2015) Epistemic petrification and the restoration of epistemic trust: A new conceptualization of Borderline Personality Disorder and its psychosocial treatment. *Journal of Personality Disorders*, 29(5), 575-609.

Fonagy P., Luyten P. (2016) A multilevel perspective on the development of Borderline Personality Disorder. *Developmental Psychopathology*, vol. 3 *Maladaptation and Psychopathology*. Doi: 10.1002/9781119125556.devpsy317.

Flom R., Johnson S. (2011) The effects of adults' affective expression and direction of visual gaze on 12-month-olds' visual preferences for an object following a 5-minute, 1-day, or 1-month delay. *British Journal of Developmental Psychology*, 29(1), 64-85.

Franks N.R., Richardson T. (2006) Teaching in tandem-running ant. *Nature*, 439, 153.

Freundlieb M. (2017) *Spontaneous visuospatial perspective-taking in humans*. Thesis submitted to Central European University – Department of Cognitive Science.

Freundlieb M., Sebanz N., Kovács A.M. (2016) When do humans spontaneously adopt another's visuospatial perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 42(3), 401–412.

Freundlieb M., Sebanz N., Kovács A.M. (2017) Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1065–1072.

Fricker E. (1994) Against gullibility. In B.K. Matilal, A. Chakrabarti (eds.) *Knowing from words* (pp. 125–161). Dordrecht, Kluwer.

Fricker E. (2006) Second-Hand Knowledge. *Philosophy and Phenomenological Research*, 73(3), 592-618.

Friedman O., Leslie A.M. (2007) The conceptual underpinning of pretense. Pretending is not 'behaving-as-if'. *Cognition*, 105, 103-124.

Futó J., Bátki A., Koós O., Fonagy P., Gergely G. (2004) *Early Social-Interactive Determinants of Later Representational and Affect Regulative Competence in Pretend Play*. Poster presented at the 14th Biennial International Conference on Infant Studies, Chicago (May).

Futó J. (2010), *The role of marked forms of caregiver-infant interactions in children's conceptual development and the formation of their representational abilities*, PhD Doctoral thesis, Debreceni Egyetem [University of Debrecen], BTK.

Futó J., Téglás E., Csibra G., Gergely G. (2010) Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition*, 117(1), 1-8.

Galef B.G., Laland K.N. (2005) Social learning in animals: empirical studies and theoretical models. *Bioscience*, 55, 489-499.

Gallese V., Keysers C., Rizzolatti G. (2004) A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 12, 493-501.

Gallese V., Lakoff G. (2005) The brain's concepts: the role of the sensory motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3–4), 455–479.

Gelman,S. A. (2003) *The essential child: Origins of essentialism in everyday thought*, Oxford UP, Oxford.

Gelman, S.A. (2004) Psychological essentialism in children. *Trends in Cognitive Science,*8(9), 404-409.

Gelman S.A., Ware E.A., Manczak E.M., Graham S.A. (2013) Children's sensitivity to the knowledge expressed in pedagogical and nonpedagogical contexts. *Developmental Psychology* 49(3), 491–504.

Gergely G. (2002) The development of understanding of self and agency. In U. Goshwami (ed.) *Handbook of childhood cognitive development* (pp. 26–46). Oxford, Blackwell.

Gergely G. (2003) The development of teleological versus mentalizing observational learning strategies in infancy. *Bullettin of the Menninger Clinic*, 67(2), 113-131.

Gergely G. (2007a) Learning "about" versus learning "from" other minds: Natural pedagogy and its implications. In: Carruthers P., Laurence S., Stich S. (eds). *The innate mind: vol. 3: foundations and the future* (pp. 170-198). Oxford: Oxford UP.

Gergely G. (2007b) The social construction of the subjective Self: The role of affect-mirroring, markedness, and ostensive communication in self-development. In: Mayes L.,Fonagy P.,Target M. (eds), *Developmental Science and Psychonalysis. Innovation and integration* (pp. 45-82) London, Karnac.

Gergely G. (2011) Kinds of Agents. The Origins of Understanding Instrumental and Communicative Agency. In: Goswami U. (ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 76-105). Oxford, Wiley-Blackwell.

Gergely G. (2013) Ostensive Communication and Cultural Learning. The Natural Pedagogy Hypothesis. In J. Metcalfe, H. S. Terrace (eds.) *Agency and Joint Attention*. Oxford scholarship. DOI 10.1093/acprof:oso/9780199988341.003.0008.

Gergely G., Watson J.S. (1996) The social biofeedback model of parental affect-mirroring. *International Journal of Psycho-Analysis*, 77, 1181–1212.

Gergely G., Watson J.S. (1999) Early socio-emotional development: Contingency perception and the social-biofeedback model. In: Rochat P. (ed.), *Early social cognition: Understanding others in the first months of life* (pp. 101-136). Mahwah, NJ: Lawrence Erlbaum.

Gergely G., Bekkering H., Kiraly I. (2002) Rational imitation in preverbal infants. *Nature*, 415, 755.

Gergely G., Csibra G. (2003) Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Science*, 7(7), 287-292.

Gergely G., Király I. (2003) *Developmental changes in observational learning of novel means between 14 and 18 months: From learning from actions to learning from agents*. Poster

presented at the 2003 Biennial Meeting of the Society for Research in Child Development (SRCD), Tampa FL, April 24-27, 2003.

Gergely G., Csibra G. (2005) The social construction of the cultural mind. Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, 6(3), 463-481.

Gergely G., Csibra G. (2006) Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: Enfield N.J., Levenson S.L. (eds.) *Roots of Human Sociality: Culture, Cognition, and Human Interaction* (pp. 229-255). Oxford, Berg Publishers.

Gergely G., Egyed K., Király I. (2007) On pedagogy. *Developmental Science*, 10(1), 139-146.

Gergely G., Unoka Z. (2008a) The development of the unreflective self. In: Busch F.N. (ed.), *Mentalization. The considerations, research findings and clinical implications* (pp. 57–102). New York-London, The Analytic Press - Taylor and Francis Group.

Gergely G., Unoka Z. (2008b) Attachment and mentalization in humans. In: Jurist E.L., Slade A., Bergner S. (eds.), *Mind to mind: Infant research, neuroscience, and psychoanalysis* (pp. 50 – 87). New York, Other Press.

Gergely G., Unoka Z. (2008/2013) Attachment, affect-regulation, and mentalization: The developmental origins of the representational affective self. In: Fonagy P., Goodyer I., *Social cognition and Developmental Psychopathology*. Oxford, Oxford UP. I refer to electronic edition. DOI: 10.1093/med/9780198569183.003.0011.

Gergely G., Koós O., Watson J.S. (2010) Contingent parental reactivity in early socio-emotional development. In: Fuchs T., Sattel H.C., Henningsen P. (eds.), *The Embodied Self. Dimensions, coherence and disorders* (pp. 141-169). Stuttgart, Schattauer GmbH.

Gliga T., Csibra G. (2007) Seeing the face through the eyes: A developmental perspective on face expertise. *Progress in Brain Research*, 164, 323-339.

Gliga T., Csibra, G. (2009) One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, 20, 347–53.

Gliga T., Volein A., Csibra G. (2010) Verbal labels modulate perceptual object processing in 1-year-old children. *Journal of Cognitive Neuroscience*, 22, 2781–2789.

Goldman A. (1999) *Knowledge in a social world*. Oxford, Clarendon Press.

Gómez J.C. (2008) The evolution of pretence: From intentional availability to intentional non-existence, *Mind&Language*, 23(5), 586–606.

Gopnik A., Wellman H.M. (1994) The theory. In: Hirschfeld, Gelman S. (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257-293). Cambridge University Press.

Grandy R., Warner R. (2013) Paul Grice. In *Stanford Encyclopedia of Philosophy* (https://plato.stanford.edu/entries/grice/#EvePsyExp)

Grice P. (1957) Meaning. *Philosophical Review*, 66, 377-388.

Grice P. (1968) Utterer's meaning, sentence meaning and word meaning. *Foundations of Language*, 4, 225-242.

Grice P. (1969) Utterer's meaning and intentions. *The Philosophical Review*, 78, 147- 177.

Grice P. (1986) Reply to Richards. In: Grandy R.E., Warner R. (eds) *Philosophical grounds of rationality: Intentions, categories, ends* (pp. 45–106). Oxford, Clarendon Press.

Grice P. (1989) *Studies in the Way of Words*, Harvard University Press, Cambridge (MA).

Griffiths P.E., Scarantino A. (2009) Emotions in the wild: The situated perspective on emotion. In: Robbins P., Aydede M. (eds.), *The Cambridge Handbook of Situated Cognition* (pp. 437-453). Cambridge, Cambridge University Press.

Grossmann T., Johnson M.H., Farroni T., Csibra G. (2008a) Social perception in the infant brain: Gamma oscillatory activity in response to eye gaze. *Social Cognitive and Affective Neuroscience*, 2, 284–291.

Grossmann T., Johnson M.H., Lloyd-Fox S., Deligianni F., Elwell C., Csibra G. (2008b) Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B*, 275, 2803-2811.

Grossmann T., Johnson M.H. (2010) Selective prefrontal cortex responses to joint attention in early infancy. *Biology Letters*, 6, 540 – 543.

Grossmann T., Parise E., Friederici A.D. (2010) The detection of communicative signals directed at the self in infant prefrontal cortex. *Frontiers in Human Neuroscience*, 4, 12 http://journal.frontiersin.org/article/10.3389/fnhum.2010.00201/full

Halberda J. (2018) Logic in babies. 12-month-olds spontaneously reason using process of elimination. *Science*, 359(6381), 1214-1215.

Hall D.G., Corrigall K., Rhemtulla M., Donegan F., Xu F. (2008) Infants' use of lexical-category-to-meaning links in object individuation. *Child Development*, 79(5), 1432-1443.

Hamilton A.F., Grafton S.T. (2006) Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, 26(4), 1133-1137.

Hamilton A.F. (2013) The mirror neuron system contributes to social responding. *Cortex*, 49, 2957-2959.

Hare B., Call J. Tomasello M. (2006) Chimpanzees deceive a human competitor by hiding. *Cognition*, 101, 495-514.

Harris P.L. (2007) Trust. *Developmental Science*, 10(1), 135–138.

Harris P.L., Kavanaugh R.D. (1993) Young children's understanding of pretence. *Monographs of the Society for Research in Child Development, 58*(1)[231], v-92.

Harris P.L., Corriveau K.H. (2011) Young children's selective trust in informants. *Philosophical Transaction of Royal Society B*, 366(1567), 1179–1187.

Harris P.L., Lane J.D. (2014) Infants understand how testimony works. *Topoi: An International Review of Philosophy*, 33(2), 443–458.

Hasan A. (2014) Knowledge by acquaintance vs. description, in *Stanford Encyclopedia of Philosophy*, http://plato.stanford.du/entries/knowledge-acquaindescrip/

He Z., Bolz M., Baillargeon R. (2011) False-belief understanding in 2.5-year-olds: evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Developmental Science*, 14(2), 292–305.

Hewlett B.S., Fouts H.N., Boyette A. H., Hewlett B.L. (2011) Social learning among Congo Basin hunter– gatherers. *Philosophical Transactions of the Royal Society B*, 366, 1168–1178.

Helming K.A., Strickland B., Jacob P. (2014) Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, 18(4), 167-170.

Helming K.A., Strickland B., Jacob P. (2016) Solving the puzzle about early belief-ascription. *Mind & Language*, 31(4), 438–469.

Hernik M., Csibra G. (2009) Functional understanding facilitates learning about tools in human children. *Current opinion in neurobiology*, 19, 34-38.

Hernik M., Southgate V. (2012) Nine-months-old infants do not need to know what the agent prefers in order to reason about its goals: on the role of preference and persistence in infants' goal-attribution. *Developmental Science*, 15(5), 714–722.

Heyes C. (2010), Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, 34, 575–583.

Heyes C.M. (2014) False belief in infancy: a fresh look. *Developmental Science*, 17(5), 647–659.

Heyes C.M. (2016) Born pupils? Natural pedagogy and cultural pedagogy. *Perspective on Psychological Science*, 11(2), 280-295.

Heyes C.M., Frith C.D. (2014) The cultural evolution of mind reading. *Science*, 344(6190), 1357–1361.

Hicók G. (2013) Do mirror neurons subserve action understanding?. *Neuroscience letters*, 540, 56-58.

Hoehl S., Reid V., Mooney J., Striano T. (2008) What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science*, 11(1), 10–16.

Hoppitt W., Brown G.R., Kendal R.L., Rendell L., Thornton A., Webster M.M., Laland K.N. (2008) Lessons from animal teaching. *Trends in Ecology & Evolution*, 23, 486–493.

Hrdy S.B. (2009) *Mothers and Others: The evolutionary origins of mutual understanding*. Cambridge, Belknap Press.

Huemer M., Perner J., Leahy B. (2017) Mental files theory of mind: When do children consider agents acquainted with different object identities? *Cognition*, 171, 122-129.

Hume D. (1748) An enquiry concerning human understanding. In: ID. (1777) *Essays concerning human understanding*. London.

Hunnius S., Bekkering H. (2014) What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transaction of the Royal Society B*, 369(1644), 1-9.

Hyde D.C., Aparicio Betancourt M., Simon C.E. (2015) Human temporal-parietal junction spontaneously tracks others' beliefs: A functional near-infrared spectroscopy study. *Human Brain Mapping*, 36(12), 4831-4846.

Iverson J. M., Capirci O., Longobardi E., Caselli M.C. (1999) Gesturing in mother–child interactions. *Cognitive Development*, 14, 57–75.

Jacob P. (2008) What do mirror neurons contribute to human social cognition? *Mind&Language*, 23(2), 190–223.

Jacob P. (2009a) The tuning-fork model of social human cognition: a critique. *Conscious and Cognition*, 18(1), 229-243.

Jacob P. (2009b) A philosopher's reflections on the discovery of mirror neurons. *Topics in Cognitive Science*, 1, 570–595.

Jacob P., Jeannerod M. (2005) The motor theory of social cognition: A critique. *Trends in Cognitive Sciences*, 9(1), 21–25.

Jacob P. (2014) Intentionality. In: Zalta E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (https://plato.stanford.edu/archives/win2014/entries/intentionality).

Jacob P. (2015) *Why reading minds is not like reading words*. Comment published on line in the following blog: http://cognitionandculture.net/blog/brent-stricklands-blog/why-reading-minds-is-not-like-reading-words.

Jacob P. (2016) *Challenging the two system account*. Paper sent me by mail by the author himself.

Jacob P., Gergely G. (2012) Reasoning about instrumental and communicative agency in human infancy. In J. B. Benson, F. Xu, T. Kushnir (eds.) *Rational Constructivism in Cognitive Development* (pp. 59–94). Elsevier, Academic Press.

Jaswal V.K., Neely L.A. (2006) Adults don't always know best. Preschoolers use past reliability over age when learning new words. *Psychological Science*, 17, 757–758.

Jaswal V.K., Croft A.C., Setia A.R., Cole C.A. (2010) Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, 21(10), 1541-1547.

Johnson M.H. (1994) Evidence against an eye direction detection module. *Current Psychology of Cognition*, 13, 630-637.

Johnson M.H. (2001) Functional brain development in humans. *Nature Reviews in Neuroscience*, 2, 475-483.

Johnson M.H. (2005) Subcortical face processing. *Nature Neuroscience Reviews*, 6, 766–77.

Johnson M.H., Morton J. (1991a) CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98, 164–181.

Johnson M.H., Morton J. (1991b) *Biology and cognitive development. The case of face recognition*, Oxford, Blackwell.

Johnson M.H., Farroni T. (2003) Perceiving and acting on eye gaze. In O. Pascalis, A. Slate (eds.) *Face processing in infancy and early childhood* (pp. 9602–9605). New York, NOVA Science Publisher.

Kaminski J., Riedel J., Call J., Tomasello M. (2005) Domestic goats (*Capra hircus*) follow gaze direction and use some cues in an object choice task. *Animal Behavior*, 69(1), 11-18.

Kelemen D. (1999) Function, goals and intention: children's teleological reasoning about objects. *Trends in Cognitive Science*, 3(12), 461-468.

Kampe K.K.W., Frith C.D., Frith U. (2003) "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing" regardless of modality. *Journal of Neuroscience*, 23, 5258–5263.

Kampis D. (2017) *Mindreadears in the crib. Cognitive mechanism for representing others' mental states in human infants*. PhD Doctoral Thesis, CEU Budapest.

Kampis D., Somogyi E., Itakura, S., Király I. (2013) Do infants bind mental states to agents?. *Cognition*, 129, 232-240.

Kampis D., Parise E., Csibra G., Kovács A.M. (2015) Neural signature for sustaining objects representations attributed to others in preverbal human infants. *Philosophical Transactions of the Royal Society B*, 282(1819), 20151683.

Kampis D., Fogd D., Kovács A.M. (2017) Nonverbal components of Theory of Mind in typical and atypical development. *Infant Behavior and Development*, 48(Pt A), 54-62.

Kaufman J., Csibra G., Johnson M.H (2003) Representing occluded objects in the human infant brain. *Proceedings of the Royal Society B*, 270(2), S140–S143.

Keates J., Graham S.A. (2008) Category markers or attributes why do labels guide infants' inductive inferences? *Psychological Science*, 19, 1287-1293.

Kinzler K.D., Dupoux E., Spelke E.S. (2007) The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 12577–12580.

Király I., Csibra G., Gergely G. (2013) Beyond rational imitation: learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology*, 116, 471-486.

Kline M.A. (2015) How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, 38, 1-71.

Kline M.A., Boyd T., Henrich J. (2013) Teaching and the life history of cultural transmission in Fijian villages. *Human nature*, 24(4), 351-74.

Knudsen B., Liszkowski U. (2012) 18-Month-Olds Predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, 17(6), 672–691.

Kobayashi C., Glover G.H., Temple E. (2007) Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia*, 45(7), 1522-1532.

Koenig M.A., Clément F., Harris P.L. (2004) Trust in testimony: Children's use of true and false statements. *Psychological Science*, 15, 694–698.

Koenig M.A., Harris P.L. (2007) The basis of epistemic trust: reliable testimony or reliable sources?. *Episteme*, 4(3), 264-284.

Koenig M.A., Cole C.A., Meyer M., Ridge K.E., Kushnir T., Gelman S.A. (2015) Reasoning about knowledge: Children's evaluations of generality and verifiability. *Cognitive Psychology*, 83, 22-39.

Koós O., Gergely G. (2001) A contingency-based approach to the etiology of "disorganized" attachment: The "flickering switch" hypothesis. *Bulletin of the Menninger Clinic*, 65(3), 397–410.

Kovács A.M. (2016) Belief Files in Theory of Mind Reasoning. *Review of Philosophy and Psychology*, 7(2), 509–527.

Kovács A.M., Téglás E., Endress A.D. (2010) The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830-1834.

Kovács A.M., Kühn S., Gergely G., Csibra G., Brass M. (2014) Are All Beliefs Equal? Implicit Belief Attributions Recruiting Core Brain Regions of Theory of Mind. *PLoS One*, 9(9), e106558. DOI: 10.1371/journal.pone.0106558

Kruger A.C., Tomasello M. (1996) Cultural learning and learning culture. In D.R. Olson, N. Torrance (eds), *Handbook of education and human development: New models of learning, teaching, and schooling* (pp. 369–387). Blackwell.

Krupenyean C., Kano F., Hirata S., Call J., Tomasello M. (2016) Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354(6308), 110-114.

Kulke L., Reiß M., Krist H., Rakoczy H. (2017) How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive development* (in press), published on-line https://doi.org/10.1016/j.cogdev.2017.09.001.

Kulke L., Rakoczy H. (2018) Implicit Theory of Mind – An overview of current replications and non-replications. *Data in Brief*, 16, 101-104.

Kupán K. (2013) *The role of human ostensive-communicative and referential cues on the social learning of human infants and dogs*, PhD Thesis, ELTE Department of Ethology.

Labinaz P. (2012) Paul Grice. *Aphex*, 6, 309-345.

Lackey J. (2006) It Takes Two to Tango: Beyond Reductionism and Non-Reductionism in the Epistemology of Testimony. In J. Lackey, E. Sosa (eds.) *The epistemology of testimony* (pp. 160-189). Oxford, Oxford UP.

Lakatos G., Soproni K., Doka A., Miklosi A. (2009) A comparative approach to dogs' (*Canis familiaris*) and human infants' comprehension of various forms of pointing gestures. *Animal Cognition*, 12(4), 621-631.

Lai C., Altavilla D., Mazza M., Tambelli R., D'Argenzio M., Guadagno A., Giusti D., Aceto P., Cecchini M. (2015) Neural correlates of face recognition in the first hours of life. *38th European Conference on Visual Perception (ECVP) 2015* Liverpool, Perception, 44(S1), p. 96.

Lancy D. F., Gaskins S., Bock, J. (eds) (2009) *The anthropology of learning in childhood*. Lanham, MD, Altamira Press.

Landrum A.R., Eaves B.S., Shafto P. (2015) Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Science*, 19(3), 109-111.

Landrum A.R., Pflaum A.D, Mills C.M. (2016) Inducing knowledgeability from niceness: Children use social features for making epistemic inferences. *Journal of Cognition and Development*, 17(5), 699-717.

Lebel C., Walker L., Leemans A., Phillips L., Beaulieu C. (2008) Microstructural maturation of the human brain from childhood to adulthood. *Neuroimage*, 40(3), 1044-55.

Lecce S., Bianco F., Demicheli P., Cavallini E. (2014) Training preschoolers on first-order false belief understanding: Transfer on advanced ToM skills and metamemory. *Child Development*, 85(6), 2404-2418.

Legare C.H., Nielsen M. (2015) Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences*, 19, 688–699.

Legare C.H., Wen N.J., Herrmann P.A., Whitehouse H. (2015) Imitative flexibility and the development of cultural learning. *Cognition*, 142, 351–361.

Legare C.H., Harris P. (2016) The ontogeny of cultural learning. *Child Development*, 87(3), 633–642.

Lehrer K. (2006) Testimony and trustworthiness. In: Lackey J., Sosa E. (eds.), *The epistemology of testimony* (pp. 145-159). Oxford, Oxford UP.

Leslie A.M. (1987) Pretence and representation: the origins of "theory of mind". *Psychological Review*, 94(4), 412-426.

Leslie A.M. (1994) Pretending and believing: issues in the theory of ToMM. *Cognition*, 50, 211-238.

Leslie A.M. (1994) ToMM, ToBy, and agency: core architecture and domain specificity. In L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 119–148), Cambridge University Press.

Leslie A.M. (2005) Developmental parallels in understanding minds and bodies. *Trends in Cognitive Science*, 9(10), 459-462.

Leslie A.M., Friedman O., German T.P. (2004) Core mechanisms in "theory of mind". *Trends in Cognitive Sciences*, 8 (12), 528–533.

Leslie A.M., German T.P., Polizzi P. (2005) Belief-desire reasoning as a process of selection. *Cognitive Psychology*, 50(1), 45-85.

Leslie S.J., Khemlani S., Glucksberg S. (2011) Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, 65(1), 15-31.

Liao Y., Acar Z.A. Makeig S., Deak G. (2015) EEG imaging of toddlers during dyadic turn-taking: Mu-rhythm modulation while producing or observing social actions. *NeuroImage*, 112, 52-60.

Little E.E., Carver L.J., Legare C.H. (2016) Cultural Variation in Triadic Infant–Caregiver Object Exploration. *Child Development*, 87(4), 1130–1145.

Liu D., Wellman H.M., Tardif T., Sabbagh M.A. (2008) Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44(2), 523–531.

Liu S., Spelke E.S. (2017) Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, 160, 35-42.

Low J., Watts J. (2013) Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, 24(3), 305-11.

Lloyd-Fox S., Blasi A., Elwell, C.E. (2010) Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience Biobehavioral Review*, 34(3), 269–284.

Lloyd-Fox S., Széplaki-Köllőd B.,Yin J., Csibra G. (2015) Are you talking to me? Neural activations in 6-month-old infants in response to being addressed during natural interactions. *Cortex*, 70, 35-48.

Lohman H., Tomasello M., Meyer S. (2005) Linguistic communication and social understainding. In Astington J.W., Baird J., (eds), *Why language matters for theory of mind* (pp. 245-265), Oxford University Press. New edition (2012) Oxford Scholarship online. DOI: 10.1093/acprof:oso/9780195159912.003.0012.

Longo M.R., Bertenthal B.I. (2006) Common coding of observation and execution of action in 9-month-old infants. *Infancy*, 10, 43–59.

Luo Y. (2011) Do 10-month-old infants understand others' false beliefs? *Cognition*, 121, 289-298.

Luo Y., Baillargeon R. (2005) Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16, 601–608.

Lyons D.E., Damrosch D.H., Lin J.K., Macris D.M., Keil F.C. (2011) The scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society B*, 366, 1158–1167.

Mandler J.M. (2003) Conceptual Categorization. In Rakison D.H., Oakes L.M. (eds.), *Early Category and Concept Development. Making sense of blooming buzzing confusion* (pp. 103-131). Oxford, Oxford UP.

Mameli M. (2001) Mindreading, mindshaping, and evolution. *Biology and Philosophy*, 16, 597–628.

Mandel D.R., Jusczyk P.W., Pisoni D.B. (1995) Infants' recognition of the sound patterns of their own names. *Psychological Science*, 6, 314–317.

Mandler J.M. (2000) Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, 1, 3-36.

Marchetti A., Sangiuliano Intra F. (2015) Mentalizzazione e tempo. La comprensione della mente attraverso le età e le relazioni. Rivista internazionale di filosofia e psicologia, 6(2), 198-211.

Mareschal D., Bremner A.J. (2009) Modeling the origins of object knowledge. In: Hood B.M., Santos L.R. (eds.) *The Origins of Object Knowledge*, Oxford, Oxford Scholarship online UP. DOI:10.1093/acprof:oso/9780199216895.003.0010.

Margolis E., Laurence S. (2005/2011) Concepts. In: Zalta E.N. (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2014/entries/concepts.

Marno H., Farroni T., Dos Santos Y.V., Ekramnia M., Nespor M., Mehler J. (2015) Can you see what I am talking about? Human speech triggers referential expectation in four-month-old infants, *Scientific Report*, 5(13594), 1-10.

Marraffa M., Meini C. (2005) *La mente sociale. Le basi cognitive della comunicazione*, Roma, Laterza.

Marraffa M., Meini C. (2015) La priorità della mentalizzazione in terza persona: implicazioni per la teoria dell'attaccamento. *Attaccamento e sistemi complessi*, 2(1), 63-84.

Marraffa M., Meini C. (2016) *L'identità personale*, Roma, Carocci.

Marraffa M., Meini C. (in press) The construction of the inner experiential space: the case of affective mentalization. [Draft sent me by the autors].

Marraffa M., Viola M. (2017) Quale mappa per il dominio delle emozioni? *Sistemi intelligenti*, 29(1), 85-108.

Marshall P.J., Meltzoff A.N. (2014) Neural mirroring systems: Exploring the EEG mu rhythm in human infancy. *Developmental Cognitive Neuroscience*, 1(2), 110-123.

Marshall P.J., Meltzoff A.N. (2014) Neural mirroring mechanism and imitation in human infants. *Philosophical Transaction of the Royal Society B*, 369(1644), 1-11.

Marshall-Pescini S., Passalacqua C., Valsecchi P., Prato-Previde E. (2010) Comment on "Differential Sensitivity to Human Communication in Dogs, Wolves, and Human Infants". *Science*, 329(142), 142-c.

Martin A., Onishi K.H., Vouloumanos A. (2012) Understanding the abstract role of speech in communication at 12 months. *Cognition*, 123(1), 50–60.

Mascaro O., Sperber D. (2009) The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, 112, 367–80.

Mattos O., Hinzen W. (2015) The linguistic root of natural pedagogy. *Frontiers in Psychology*, 6(1424), 1-12.

Mazzarella D. (2016) Pragmatics, Modularity, and Epistemic Vigilance. *Argumenta*, 1(2), 181-193.

McDowell J. (2010) What is the content of an intention in action?. *Ratio*, XXIII, 415-432.

Medin D.L., Ortnoy A. (1989) Psychological essentialism. In: Vosnadiou S., Ortony A. (eds.) *Similarity and Analogical Reasoning* (pp. 179-195). Cambridge, Cambridge UP.

Meinhardt J., Sodian B., Thoermer C., Döhnel K., Sommer M.(2011) True- and false-belief reasoning in children and adults: An event-related potential study of theory of mind. *Developmental Cognitive Neuroscience*, 1, 67-76.

Meini C. (2016) La natura del senso comune in psicologia. In: Lavazza A., M. Marraffa M. (eds.) *La Guerra dei mondi. Scienza e senso comune* (pp. 183-202). Torino, Codice edizioni.

Meini C. (2017) Guardarsi dentro. Il ruolo della relazione nella formazione della coscienza introspettiva. *Sistemi intelligenti*, 29(1), 109-128.

Meini C., Paternoster A. (2012) Mirror neurons as a conceptual mechanism? *Mind & Society*, 11(2), 183-201.

Meltzoff, A.N. (1988) Infant imitation after a 1-week delay: Long term memory for novel acts and multiple stimuli. *Developmental Psychology,* 24, 470-476.

Meltzoff A.N. (1999) Origins of theory of mind, cognition and communication. *Journal of Communication Disorders*, 32, 251–269.

Meltzoff A. N. (2002) Imitation as a Mechanism of Social Cognition: Origins of Empathy, Theory of Mind, and the Representation of Action. In: Goswami U. (ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 6.25). Oxford, Wiley-Blackwell.

Meltzoff A. N. (2005) Imitation and other minds: The like-me hypothesis. In: Hurley S., Chater N. (eds.), *Perspectives on imitation: From neuroscience to social science*, Cambridge, MIT Press, 55-77.

Meltzoff A.N. (2007a) 'Like me': a foundation for social cognition. *Developmental Science*, 10(1), 126-132.

Meltzoff A.N. (2007b) The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychologica*, 124(1), 26–43.

Mercier H., Sperber D. (2011) Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 35, 57–74.

Michael J., Christensen W. (2015) Flexible goal attribution in early mindreading. *Psychological Review*, 123(2), 219-227.

Michael J. (2015) Putting unicepts to work: a teleosemantic perspective on the infant mindreading puzzle. *Synthese*. DOI: 10.1007/s11229-015-0850-x.

Michaelian K. (2010) In defence of gullibility: the epistemology of testimony and the psychology of deception detection. *Synthese*, 176, 399–427.

Michaelian K. (2012a) Metacognition and endorsement. *Mind & Language*, 27(3), 284–307.

Michaelian K. (2012b) (Social) Metacognition and (Self-)Trust. *Review of Philosophy and Psychology*, 3, 481–514.

Mikhalevich I., Powell R., Logan C. (2017) Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus*, 7(3): 20160121.

Millikan R.G. (1984) *Language, thought, and other biological categories*, Cambridge, MIT Press.

Millikan R.G. (2000) *On clear and confused ideas. An essay about substance concepts*. Cambridge, Cambridge UP.

Millikan R. (2013) An epistemology for phenomenology? In: Brown R. (ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience*. Springer's series Studies in brain and mind, 6 (pp. 13–26). Netherlands, Springer.

Moll H. (2017) The Transformative Cultural Intelligence Hypothesis: Evidence from Young Children's Problem-Solving. *Review of Philosophy and Psychology*, 1-15. (published on line)

Montgomery D.E. (2005) The Developmental Origins of Meaning for Mental Terms. In Astington J.W., Baird J., (eds) *Why language matters for theory of mind* (pp. 106-122), Oxford UP. New edition (2012) Oxford Scholarship online. DOI:10.1093/acprof:oso/9780195159912.003.0006

Moore C., Dunham P.J. (Eds.) (1995) *Joint Attention. Its origins and role in development*. Hillsdale, Lawrence Erlbaum.

Moore R. (2013) Social learning and teaching in chimpanzees. *Biology and Philosophy*, 28 (6), 879-901.

Moore R. (2014) The ontogenetic constraints on Grice's theory of communication. In Matthews D. (ed) *Pragmatic development in first language acquisition* (pp. 87-104). Amsterdam, John Benjamins publishing company.

Moore R. (2016a) Pedagogy and social learning in human development. In Kiverstein J. (ed.) *The Routledge Handbook of Philosophy of the Social Mind* (pp. 35-52). London, Routledge.

Moore R. (2016b) Gricean communication and cognitive development. *The Philosophical Quarterly*, 67(267), 303–326.

Moore R. (in press) Non-Natural Pedagogy.

Moses L., Baldwin D., Rosicky J., Tidball G. (2001) Referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, 72, 718–735.

Möttönen R., Farmer H., Watkins K.E. (2016) Neural basis of understanding communicative actions: Changes associated with knowing the actor's intention and the meanings of the actions. *Neuropsychologia*, 81, 230-237.

Mumme D.L., Fernald A. (2003) The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*, 74(1), 221–237.

Murray L., Trevarthen C. (1985) Emotion regulation of interactions between two-month-olds and their mother. In: Field T.M., Fox N.A. (eds.) *Social perception in infants* (pp. 177-198). Norwood, Ablex.

Murray L., De Pascalis L., Bozicevic L., Hawkins L., Sclafani V., Ferrari P.F. (2016) The functional architecture of mother-infant communication, and the development of infant social expressiveness in the first two months. *Scientific Reports*, 6, 1-9. DOI: 10.1038/srep39019.

Nagy E. (2008) Innate intersubjectivity: Newborns' sensitivity to communication disturbance. *Developmental Psychology*, 44, 1779–1784.

Nakao H., Andrews K. (2014) Ready to teach or ready to learn: A critique of the natural pedagogy theory. *Review of Philosophy and Psychology*, 5(4), 465–483.

Nelson C.A., Snyder K. (2005) The segregation of face and object processing in development: A model system of categorization?. In Gershkoff-Stowe L., Rakison D. (eds.), *Building object categories in developmental time* (pp. 1-31). Lawrence Erlbaum.

Nielsen M. (2013) Young children's imitative and innovative behaviour on the floating object task. *Infant and Child Development*, 22(1), 44-52.

Nielsen M., Tomaselli K. (2010) Overimitation in Kalahari bushman children and the origins of human cultural cognition. *Psychological Science*, 21, 729–736.

Oktay-Gür N., Schulz A., Rakoczy H. (2018) Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, 173, 60-74.

Onishi K.H., Baillargeon R. (2005) Do 15-month-olds understand false beliefs? *Science*, 308, 255-258.

Onishi K.H., Baillargeon R., Leslie A. M. (2007) 15-Month-old infants detect violations in pretend scenarios. *Acta Psychologica*, 124(1), 106–128.

Origgi G. (2004) Is trust an epistemological notion? *Episteme*, 1(1), 61-72.

Over H., Carpenter M. (2012) Putting the social into social learning: explaining both selectivity and fidelity in children's copying behaviour. *Journal of comparative psychology*, 126(2), 182-92.

Over H., Carpenter M. (2013) The social side of imitation. Child Development Perspectives, 7, 6-11.

Pagin P. (2013) The cognitive significance of Mental Files, *Disputatio*, 5(36), 133-145.

Palmquist C.M., Jaswal V.K. (2015) Preschoolers' inferences about pointers and labelers: The modality matters. *Cognitive Development*, 15, 178-185.

Palmquist C.M., Jaswal V.K., Rutherford A. (2016) Success inhibits preschoolers' ability to establish selective trust. *Journal of Experimental Child Psychology*, 152, 192-204.

Pannu J.K., Kaszniak A.W. (2005) Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, 15(3), 105-130.

Panksepp J. (1998) *Affective Neuroscience*. Oxford: Oxford University Press.

Parise E., Reid V.M., Stets M., Striano T. (2008) Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, 3(2), 141-150.

Parise E., Friederici A.D., Striano T. (2010) ''Did you call me?'' 5-month-old infants own name guides their attention. *PLoS One*, 5, e14208.

Parise E., Csibra G. (2013) Neural responses to multimodal ostensive signals in 5-monthold infants. *PloS one*, 8, e72360.

Pascalis O., de Schonen S., Morton J. (1995) Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, 18, 79–85.

Paulus M., Hunnius S., Vissers M. Bekkering H. (2011a) Imitation in infancy: rational or motor resonance? *Child Development*, 82(4), 1047–1057.

Paulus M., Hunnius S., Vissers M., Bekkering H. (2011b) Bridging the gap between the other and me: The functional role of motor resonance and action effects in infants' imitation. *Developmental Science*, 14, 901–910.

Pellegrini A.D., Smith P.K. (2004) *The Nature of Play: Great Apes and Humans*. New York, Guilford Press.

Perconti P. (2011) *Coscienza*. Bologna, Il Mulino.

Perconti P. (2015) *La prova del budino. Il senso comune e la nuova scienza della mente*. Milano, Mondadori.

Perconti P. (2017) *Filosofia della mente*. Bologna, Il Mulino.

Perner J. (1991) *Understanding the representational mind*, Cambridge, MIT Press.

Perner J., Ruffman T. (2005) Infants' insight into the mind: How deep?. *Science*, 308, 214-216.

Perner J., Roessler J. (2012) From infants' to children's appreciation of belief. *Trends in Cognitive Sciences*, 16(10), 519-525.

Perner J., Huemer M., Leahy B. (2015) Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition*, 145, 77-88.

Perner J., Leahy B. (2016) Mental Files in Development: Dual Naming, False Belief, Identity and Intensionality. *Review of Philosophy and Psychology*, 7 (2), 491-508.

Pesch A., Suárez S., Koenig M.A. (2018) Trusting others: shared reality in testimonial learning. *Current Opinion in Psychology*, 23, 38-41.

Peterson C.C., Wellman H.M., Liu D. (2005) Steps in theory-of-mind development for children with deafness or autism. *Child Development*, 76(2), 502-517.

Peterson C.C. (2009) Development of social-cognitive and communication skills in children born deaf. *Scandinavian Journal of Psychology*, 50(5), 475–83.

Phillips J., Ong D.C., Surtees A.D.R., Xin Y., Williams S., Saxe R., Frank M.C. (2015) New look at automatic Theory of Mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*. Published on-line. DOI: 10.1177/0956797614558717.

Piaget J. (1954) *The Construction of Reality in the Child*. New York, Basic Books.

Pineda J.A. (2005) The functional significance of mu rhythms: translating ‘‘seeing’’ and ‘‘hearing’’ into ‘‘doing’’. *Brain Research Review*, 50, 57– 68.

Pineda J.A., Hecht E. (2009) Mirroring and mu rhythm involvement in social cognition: Are there dissociable subcomponents of theory of mind?. *Biological Psychology*, 80(3), 306-314.

Pinkham A.M., Jaswal V.K. (2011) Watch and learn? Infants privilege efficiency over pedagogy during imitative learning. *Infancy*, 16(5), 535–544.

Pomiechowska B., Csibra G. (2017) Motor activation during action perception depends on action interpretation. *Neuropsychologia*, 1-28. DOI: 10.1016/j.neuropsychologia.2017.01.032.

Poulin-Dubois D., Brooker I., Polonia A. (2011) Infants prefer to imitate a reliable person. *Infant Behavior Development*, 34, 303–309.

Poulin-Dubois D., Brosseau-Liard P. (2016) The developmental origins of selective social learning. *Current Directions in Psychological Science*, 25(1), 60-64.

Powell L.J., Hobbes K., Bardis A., Carey S., Saxe R. (2017) Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, published on line http://dx.doi.org/10.1016/j.cogdev.2017.10.004.

Prasada S. (2000). Acquiring generic knowledge. *Trends in Cognitive Science*, 4, 66–72.

Premack D. (1990) The infant’s theory of self-propelled objects. *Cognition*, 36, 1–16.

Premack D., Woodruff G. (1978) Does the chimpanzee have a “Theory of mind”?. *Behavioral and Brain Science*, 4, 515-526.

Premack D., Premack, A.J. (1995) Intention as psychological cause. In: Sperber D., Premack D., Premack A.J. (eds.) *Causal cognition: A Multidisciplinary debate* (pp. 185–199). Clarendon Press.

Priewasser B., Rafetseder E., Gargitter C., Perner J. (2017) Helping as an early indicator of a theory of mind: Mentalism or teleology? *Cognitive Development*, http://dx.doi.org/10.1016/j.cogdev.2017.08.002 R

Prinz W. (2006) What re-enactment learns us. *Cortex, 42*, 515-517.

Rakoczy H. (2012) Do infants have a theory of mind? *British Journal of Developmental Psychology*, 30(1), 59–74.

Rakoczy H., Bergfeld D., Schwarz I., Fizke E. (2015) Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development*, 86(2), 486-502.

Raihani N.J., Ridley A.R. (2008) Experimental evidence for teaching in wild pied babblers. *Animal behavior*, 74, 1303-1309.

Range F., Virnayi Z., Huber L. (2007). Selective imitation in domestic dogs. *Current Biology*, 17, 868– 872.

Rayson H., Bonaiuto J.J., Ferrari P.F., Murray L. (2016) Mu desynchronization during observation and execution of facial expressions in 30-month-old children. *Developmental Cognitive Neuroscience*, 19, 279-287.

Recanati F. (1997) Can we believe what we do not understand? *Mind&Language*, 12, 84–100.

Recanati F. (2011) Mental files and identity. *Anne Reboul. Philosophical Papers Dedicated to Kevin Mulligan*, University of Geneva, electronic publication, 1-26.

Recanati F. (2012) *Mental files*, Oxford, Oxford UP.

Redcay E., Saxe R. (2013) Do you see what I see? The neural bases of joint attention. In J. Metcalfe, H. S. Terrace (eds.) *Agency and Joint Attention*. Oxford scholarship. DOI:10.1093/acprof:oso/9780199988341.003.0014.

Reid T. (1764/1997) *An inquiry into the human mind on the principles of common sense*. In Brookes D.R. (ed.), University Park, Pennsylvania State UP.

Reid T. (1788/1983) *Essay on the intellectual powers of man*. In Beanblossom R., Lehrer R. (eds) *Thomas Reid's Inquiry and Essays*. Indianapolis, Hackett.

Reid V.M., Striano T. (2005) Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience. *European Journal of Neuroscience*, 21, 1763-1766.

Repacholi B.M. (1998) Infants' use of attentional cues to identify the referent of another person's emotional expression. *Developmental Psychology*, 34, 1017–1025.

Repacholi B.M., Gopnik A. (1997) Early reasoning about desires: Evidence from 14- and 18months-olds. *Developmental Psychology*, 33(1), 12-21.

Richardson H., Lisandrelli G., Riobueno-Naylor A., Saxe R. (2018) Development of the social brain from age three to twelve years. *Nature communication*, 9(1027). DOI: 10.1038/s41467-018-03399-2.

Rizzolatti G., Fogassi L., Gallese V. (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Review Neuroscience*, 2, 661–670.

Rizzolatti G., Craighero L. (2004) The mirror-neuron system. *Annual Review Neuroscience*, 27, 169-192.

Rizzolatti G., Fogassi L. (2014) The mirror mechanism: recent findings and perspectives. *Philosophical Transaction of the Royal Society B*, 369(1644), 1-4.

Ronfard S., Bartz D.T., Cheng L., Chen X., Harris P.L. (2018) Children's developing ideas about knowledge and its acquisition. *Advances in Child Development*, (in press) available on-line https://doi.org/10.1016/bs.acdb.2017.10.005

Rubio-Fernández P. Geurts B. (2013) How to pass the false-belief task before your fourth birthday. *Psychological Science*, 24(1), 27-33.

Rupert R. (2008) Causal theories of mental content. *Philosophy Compass*, 3(2), 353–380.

Russell J.A. (2003), Core affect and the psychological construction of emotion. *Psychological review*, 110(1), 145-172.

Sabbagh M.A., Xu F., Carlson S.M., Moses L.J., Lee K. (2006) The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychological Science*, 17(1):74-81.

Sabbagh M.A., Bowman L.C., Evraire L.E., Ito J.M. (2009) Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*, 80(4), 1147-62.
Sage K., Baldwin D. (2010) Social gating and pedagogy: Mechanisms for learning and implications for robotics. *Neural Networks*, 23, 1091-1098.

Sage K.D., Baldwin D. (2011) Disentangling the social and the pedagogical in infants' learning about tool-uses. *Social Development*, 20(4), 825-844.

Sage K.D., Baldwin D. (2012) Exploring natural pedagogy in play with preschoolers: Cues parents use and relations among them. *Education Research and Perspectives*, 39, 153-181.

Samson D., Apperly, I.A., Braithwaite J.J., Andrews B.J., Bodley Scott S.E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1255–1266.

Sartori L., Betti S., Castiello U. (2013) When mirroring is not enough: that is, when only a complementary action will do (the trick). *Neuroreport*, 24, 601-604.

Saxe R., Wexler A. (2005) Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, 43(10), 1391-1399.

Scarantino A. (2012) Discrete emotions: From folk psychology to causal mechanisms. In: P. Zachar P., Ellis R. (eds.), *Categorical and Dimensional Models of Affect: A seminar on the theories of Panksepp and Russell* (pp. 135-154). Amsterdam, John Benjamins.

Scarantino A. (2015) Basic emotions, psychological construction, and the problem of variability. In: Barrett L.F., Russell J.A. (eds.), *The Psychological Construction of Emotion* (pp. 334-76). New York, Guilford Press.

Scarantino A., Griffiths P. (2011) Don't give up on basic emotions. *Emotion Review*, 3(4), 444-454.

Scheel M.H., Shaw H.L., Allen Gardner R. (2015) More examples of chimpanzees teaching. *Behavioral and Brain Sciences*, 38, 44.

Schmidt M.F.H., Rakoczy H., Tomasello M. (2011) Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science*, 14(3), 530-539.

Scholl B.J., Leslie A.M. (1999) Modularity, Development and Theory of Mind. *Mind & Language*, 14(1), 131-153.

Schwier C., van Maanen C., Carpenter M., Tomasello M. (2006) Rational imitation in 12-month-old infants. *Infants*, 10(3), 303–311.

Schwitzgebel E. (2015) Belief. In Zalta E.N. (ed.) *The Stanford Encyclopedia of Philosophy (*https://plato.stanford.edu/archives/sum2015/entries/belief).

Scott R.M. (2017) Surprise! 20-month-old infants understand the emotional consequences of false beliefs. *Cognition*, 159, 33-47.

Scott L.S., Nelson C.A. (2004) The developmental neurobiology of face processing. In J.M. Oldman, M.B. Riba (Eds.), *Review of psychiatry series*, vol. 23 (pp. 29-68). American Psychiatric Publishing, Inc.

Scott R.M., Baillargeon R. (2009) Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, 80, 1172–1196.

Scott R.M., Baillargeon R., Song H., Leslie A.M. (2010) Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, 61, 366–395.

Scott R.M., He Z., Baillargeon R., Cummins D. (2012) False-belief understanding in 2.5-year-olds: evidence from two novel verbal spontaneous-response tasks. *Developmental Science*, 15(2), 181-193.

Scott R.M., Baillargeon R. (2013) Do infants really expect agents to act efficiently? A critical test of the Rationality Principle. *Psychological Science*, 24(4), 466-474.

Scott R.M., Richman J.C., Baillargeon R. (2015) Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, 82, 32–56.

Senju A., Johnson M.H., Csibra G. (2006) The development and neural basis of referential gaze perception. *Social Neuroscience*, 1(3-4), 220-234.

Senju A., Csibra G. (2006) Gaze following in human infants depends on communicative signals. Current Biology, 18, 668-671.

Senju A., Csibra G., Johnson M.H. (2008) Understanding the referential nature of looking: Infant's preference for object-direct gaze. *Cognition*, 108, 303-319.

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Scence*, 13(3), 127–134.

Senju A, Southgate V, Miura Y, Matsui T, Hasegawa T, Tojo Y, Osanai H, Csibra G. (2010) Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Developmental Psychopathology*, 22(2), 353-360.

Senju A., Southgate V., Snape C., Leonard M., Csibra G. (2011) Do 18-months-olds really attribute mental states to others? A critical test. *Psychological Science*, 22(3), 878-880.

Shafto P., Goodman N. (2008) Teaching games: Statistical sampling assumptions for pedagogical situations. *Proceedings of the 30th annual conference of the Cognitive Science Society*.

Shneidman L., Woodward A.L. (2015) Are Child-Directed Interactions the Cradle of Social Learning?. *Psychological Bulletin*, 142(1), 1-17.

Shneidman L. Gaskins S., Woodward A.L. (2015) Child-directed teaching and social learning at 18 months of age: evidence from Yucatec Mayan and US infants. *Developmental Science*, 19(3), 372-381.

Shneidman L., Gweon H., Sculz L.E., Woodward A.L. (2016) Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development*, 87(3), 723–735.

Simpson E.A., Murray L., Paukner A., Ferrari P.F. (2014) The mirror neuron system as revealed through neonatal imitation: presence from birth, predictive power and evidence of plasticity. *Philosophical Transaction of the Royal Society B*, 369(1644), 1-12.

Sinclair R. (2005) The philosophical significance of triangulation: Locating Davidson's non-reductive naturalism. *Metaphilosophy*, 36, 708–27.

Sinigaglia C. (2010) Mirroring and understanding action. In: Suárez M., Dorato M., Rédei M. (eds.) *EPSA Philosophical Issues in the Sciences* ((pp. 227-238), Dordrecht, Springer.

Sinigaglia C., Rizzolatti G. (2011) Through the looking glass: Self and others. *Consciousness and cognition*, 20(1), 64-74.

Sobel D., Sommerville J. (2009) Rationales in children's causal learning from others' actions. *Cognitive Development*, 24(1), 70-79.

Sommer M., Döhnel K., Sodian B., Meinhardt J., Thoermer C.; Hajak G. (2007) Neural correlates of true and false belief reasoning. *Neuroimage*, 35(3), 1378–1384.

Song H., Onishi K., Baillargeon R., Fisher C. (2008) Can an actor's false belief be corrected by an appropriate communication? Psychological reasoning in 18.5-month-old infants. *Cognition*, 109(3), 295–315.

Song H., Baillargeon R. (2008) Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44(6), 1789–1795.

Southgate V. (2013) Early manifestations of mindreading. In: S. Baron-Cohen, M. Lombardo, H. Tager-Flusberg, *Understanding Other Minds: Perspectives from developmental social neuroscience*, Oxford, Oxford Scholarship Online. My quotations refer to the online version available. DOI:10.1093/acprof:oso/9780199692972.003.0001

Southgate V., Senju A., Csibra G. (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587–592.

Southgate V., Johnson M.H., Csibra G. (2008) Infants attribute goals even to biomechanically impossible actions. *Cognition*, 107(3), 1059-1069.

Southgate V., Johnson M.H., Tamsin Osborne T., Csibra G. (2009) Predictive motor activation during action observation in human infants. *Biology Letters. Animal Behavior*, 5(6), 769–772.

Southgate V., Chevallier C., Csibra G. (2010a) Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, 12(6), 1013–1019.

Southgate V., Chevallier C., Csibra G. (2010b) Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13(6), 907–912.

Southgate V., Begus K. (2013) Motor activation during the prediction of nonexecutable actions in infants. *Psychological Science*, 24(6), 828-835.

Southgate V., Begus K., Lloyd-Fox S., di Gangi V., Hamilton A. (2014) Goal representation in the infant brain. *Neuroimage*, 85, 294-301.

Southgate V., Vernetti A. (2014) Belief-based action prediction in preverbal infants. *Cognition*. 130(1), 1-10.

Spaulding S. (2013) Commentary on 'How to construct a minimal theory of mind'. http://philosophyofbrains.com/2013/11/11/symposium-on-butterfill-and-apperlys-how-to-construct-a-minimal-theory-of-mind-mind-language-28-5-606-63.aspx

Sperber D. (1997) Intuitive and reflective beliefs. *Mind and Language*, 12, 67–83.

Sperber D. (2000) Metarepresentations in an evolutionary perspective. In: Sperber D. (ed) *Metarepresentations*: *A Multidisciplinary Perspective* (pp. 117-137). Oxford UP.

Sperber D. (2001) An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29, 401–413.

Sperber D. (2005) Modularity and relevance. How can a massively modular mind be flexible and context sensitive? In: Carruthers P., Laurence S., Stich S. (eds). *The innate mind: vol. 1: Structure and Content* (pp. 53-68). Oxford: Oxford UP.

Sperber D., Wilson D. (1986/1995) *Relevance: Communication and Cognition*, Second edition, Oxford, Blackwell.

Sperber D., Wilson D. (2002) Pragmatics, Modularity and Mind-reading. *Mind & Language*, 17, 3-23.

Sperber D., Hirschfeld L. (2004) The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, 8(1), 40-46.

Sperber D., Clément F., Heintz C., Mascaro O., Mercier H., Origgi G., Wilson D. (2010) Epistemic vigilance. *Mind & Language*, 25(4), 359–393.

Sprung M., Perner J., Mitchell P. (2007) Opacity and embedded perspectives: Object identity and object properties. *Mind & Language*, 22, 215-245.

Stalnaker R. (2002) Common ground. *Linguistics and Philosophy*, 25, 701-21.

Steglich-Petersen A., Michael J. (2015) Why desire reasoning is developmentally prior to belief reasoning. *Mind & Language*, 30(5), 526-549.

Stephens E.C., Koenig M.A. (2015) Varieties of testimony: Children's selective learning in semantic versus episodic domains. *Cognition*, 137, 182-188.

Sterelny K. (1991) *The Representational Theory of Mind. An Introduction*, Cambridge, Basil Blackwell.

Sterelny K. (2012) *The evolved apprentice. How evolution makes human unique*, Cambridge, MIT PRess

Stich S. (1978) Beliefs and subdoxastic states. *Philosophy of Science*, 45, 499–518.
Strauss S., Ziv M. (2012) Teaching is a natural cognitive ability among humans. *Mind, Brain and Education*, 6(4), 186-196.

Striano T., Stahl D. (2005) Sensitivity to triadic attention in early infancy. *Developmental Science*, 8, 333-343.

Striano T., Chen X., Cleveland A., Bradshaw S. (2006a) Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, 3(3), 289-299.

Striano, T., Reid, V.M., Hoehl S. (2006b) Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience*, 23, 2819-2823.

Surian L., Caldi S., Sperber D. (2007) Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580-586.

Surtees A., Apperly I.A., Samson D. (2013) Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, 129, 426– 438.

Tanone D., Csibra G. (2015) Learning in and about opaque worlds. *Behavioral and Brain Sciences*, 38, 49-50.

Tauzin T. (2017) *The comparative investigation of the pedagogical stance. The role of ostensive-communicative cues in social learning in dogs and human infants*. PhD Doctoral Thesis, University of Technology and Economics PhD School in Psychology – Cognitive Science, Budapest.

Tauzin T., Gergely G. (2018) Communicative mind-reading in preverbal infants. *Scientific Reports*, 8(9534), 1-9.

Thompson J.R. (2014) Meaning and Mindreading. *Mind & Language*, 29(2), 167-200.

Thornton A., Raihani N.J., Radford A.N. (2007) Teaching in the wild: some clarification. *Trends in cognitive science*, 11(7), 272-273.

Thornton A., Raihani N.J. (2008) The evolution of teaching. *Animal Behaviour*, 75, 1823-1836.

Thornton A., McAuliffe K. (2012) Teaching can teach us a lot. *Animal Behaviour*, 83(4), e6–e9.

Tirassa M., Bosco F.M., Colle L. (2006) Rethinking the ontogeny of mindreading. *Consciousness & Cognition*, 15(1), 197-217.

Tomasello M. (1995) Joint attention as social cognition. In C. Moore, P. Dunham (Eds.), *Joint attention: Its origin and role in development* (pp. 103-130). Hillsdale, Lawrence Erlbaum.

Tomasello, M. (1996). Do apes ape? In C. M. Heyes, B.G. Galef (eds.), *Social Learning in Animals: The Roots of Culture* (pp. 319-346). New York, Academic Press

Tomasello, M. (1999) *The cultural origins of human cognition*. Harvard University Press.

Tomasello M. (2008) *Origins of human communication*. Cambridge, MIT Press.

Tomasello M. (2016) Cultural learning redux. *Child Development*, 87(3), 643–653.

Tomasello M., Kruger, A.C., Ratner, H.H. (1993) Cultural learning. *Behavioral and Brain Sciences*, 16, 495–552.

Tomasello M., Call J., Hare B. (2003) Chimpanzees understand psychological states - the question is which ones and to what extent. *Trends in Cognitive Sciences*, 7, 153–160.

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(5), 675-691.

Topál J., Gergely G., Miklosi A., Erdohegyi A., Csibra G. (2008) Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science*, 321, 1831-1834.

Topál J., Gergely G., Erdőhegyi Á., Csibra G., Miklósi Á. (2009) Differential sensitivity to human communication in dogs, wolves, and human infants. *Science*, 325(5945), 1269–1272.

Träuble B., Marinović V., Pauen S. (2010) Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, 15, 434–444.

Trevarthen C. (1994) The self born in intersubjectivity: The psychology of an infant communicating. In: Neisser U. (ed.), *The perceived Self: Ecological and interpersonal sources of Self Knowledge* (pp. 121-173). Cambridge, Cambridge University Press.

Trevathen C., Aitken K.J. (2001) Infant intersubjectivity: research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry, and allied disciplines*, 42(1), 3-48.

Tronick E., Als H., Adamson L., Wise S., Brazelton T.B. (1978) The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child & Adolescent Psychiatry*, 17, 1–13.

Van de Vondervoort J.W., Friedman O. (2017) Young children protest and correct pretense that contradicts their general knowledge. *Cognitive Development*, 43, 182-189.

Van Overwalle F., Vanderkerchhove M. (2013) Implicit and explicit social mentalizing: dual processes driven by a shared neural network. *Frontiers in human Neuroscience*, 7(560), doi: 10.3389/fnhum.2013.00560.

Vanderwert R.E., Fox N.A., Ferrari P.F. (2013) The mirror mechanism and mu rhythm in social development. *Neuroscience letters*, 540, 15-20.

Vannuscorps G., Caramazza A. (2016) Typical action perception and interpretation without motor simulation. *Proceedings of the National Academy of Sciences*, 113, 86-91.

Varela, F.J., Thompson E., Rosch E. (1991) *The embodied mind: Cognitive science and human experience*, Cambridge, MIT Press.

Vivanti G., Rogers S.J. (2014) Autism and the mirror neuron system: insights from learning and teaching. *Philosophical Transaction of the Royal Society B*, 369(1644), 1-8.

Vorms M. (2012), A-not-B errors: testing the limits of natural pedagogy. *Philosophy and Psychology*, 3(4), 525-545.

Vouloumanos A., Onishi K.H., Pogue A. (2012) Twelve-month-old infants recognize that speech can communicate unobservable intentions. *PNAS*, 109, 12933–12937.

Vouloumanos A., Onishi K.H. (2013) Understanding the structure of communicative interactions in infancy. In: Metcalfe J., Terrace H.S. (eds.) *Agency and Joint Attention*. Oxford scholarship. DOI 10.1093/acprof:oso/9780199988341.003.0010.

Vouloumanos A., Martin A., Onishi K.H. (2014) Do 6-month-olds understand that speech can communicate?. *Developmental Science*, 17(6), 872-879.

Wang L., Leslie A.M. (2016) Is implicit Theory of Mind the 'Real Deal'? The own-belief/true-belief default in adults and young preschoolers. *Mind & Language*, 31(2), 147–176.

Watson J.S. (1994) Detection of self: The perfect algorithm. In: Parker S., Mitchell R., Boccia M. (eds.), *Self-Awareness in animals and humans: Developmental perspectives* (pp. 131-149). NewYork, Cambridge University Press.

Webb S. J., Long J. D., Nelson C.A. (2005) A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*, 8, 605-616.

Wellman H.M. (1990) *Children's theories of mind*. Bradford, MIT Press.

Wellman H.M., Cross D., Watson J. (2001) Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684.

Wellman H.M., Liu D. (2004) Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523-541.

Wellman H.M., Peterson C.C. (2013) Theory of mind, development, and deafness. In: S. Baron-Cohen, M. Lombardo, H. Tager-Flusberg, *Understanding Other Minds: Perspectives from developmental social neuroscience*, Oxford, Oxford Scholarship Online. My quotations refer to the online version available. DOI:10.1093/acprof:oso/9780199692972.003.0004.

Westra E. (2016a) Pragmatic development and the false belief task. *Review of Philosophy and Psychology*, published on line (DOI: 10.1007/s13164-016-0320-5).

Westra, E. (2016b) Spontaneous mindreading: a problem for the two-systems account. *Synthese*, published on line (DOI :10.1007/s11229-016-1159-0).

Westra, E., Carruthers, P. (2017) Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, 158, 165-176.

Widen S.H. (2013) Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, 5(1), 72-77.

Wilcox T., Bortfeld H., Woods R., Wruck E. (2005) Using near-infrared spectroscopy to assess neural activation during object processing in infants. *Journal of Biomedical Optics*, 10(1): 11010.

Wilson D. (2000) Metarepresentation in Linguistic Communication. In: Sperber D. (ed.) *Metarepresentations*: *A Multidisciplinary Perspective* (pp. 412-448). Oxford UP.

Wilson D. (2005) New directions for research on pragmatics and modularity. *Lingua,* 115 (8), 1129 - 1146.

Wilson D., Sperber D. (2002) Relevance Theory. In L. Horn and G. Ward (eds.) *Handbook of Pragmatics* (pp. 607-632). Oxford, Blackwell.

Wilson M., Lancaster J., Emmorey K. (2010) Representational momentum for the human body: Awkwardness matters, experience does not. *Cognition*, 116(2), 242–250.

Wilson D.S. (2013) Human cultures are primarily adaptive at the group level. *Cliodynamics*, 4, 102-138. I refer to the on-line version https://evolution-institute.org/focus-article/david-sloan-wilson-human-cultures-are-primarily-adaptive-at-the-group-level/.

Wimmer H., Perner J. (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.

Wittgenstein L. (1949/1969) *On certainty*, Oxford, Basil Blackwell.

Woodward A. (1998) Infants selectively encoded the goal object of an actor's reach. *Cognition*, 69, 1-34.

Woodward A., Gerson (2014) Mirroring and the development of action understanding. *Philosophical Transaction of the Royal Society B*, 369(1644), 1-8.

Wurm M.F., Schubotz R. I. (2016) What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin & Review*, 1-7.

Xu F. (2002) The role of language in acquiring object kind concepts in infancy. *Cognition*, 85(3), 223-250.

Xu F. (2005) Categories, kinds, and object individuation in infancy. In Gershkoff-Stowe L., Rakison D. (eds.), *Building object categories in developmental time* (pp. 63-89). Lawrence Erlbaum.

Xu F., Cote M., Baker A. (2005) Labeling guides object individuation in 12-month-old infants. *Psychological Science*, 16(5), 372-377.

Yato Y., Kawai M., Negayama K., Sogon S., Tomiwa K., Yamamoto H. (2008) Infant responses to maternal still-face at 4 and 9 months. *Infant Behavior and Development*, 31, 570–577.

Yi X., Low W.Q. (2018) Willingness to revise own testimony: 3- and 4-year-olds' selective trust in unexpected testimony from accurate and inaccurate informants. *Journal of Experimental Psychology*, 17, 1-15.

Yoon J.M., Johnson M.H., Csibra G. (2008) Communication induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, 105, 13690-13695.

Zaki J., Wager T.D., Singer T., Keysers C., Gazzola V. (2016) The anatomy of suffering: understanding the relationship between nociceptive and empathic pain. *Trends in Cognitive Science*, 20, 249–259.

Zawidzki T.W. (2011) How to Interpret Infant Socio-Cognitive Competence. *Review of Philosophy and Psychology*, 2, 483–497.

Zawidzki, T.W. (2013) *Mindshaping. A new framework for understanding human social cognition*. Cambridge, MIT Press.

Ziv M., Frye D. (2004) Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development*, 19, 457–477.

Ziv M., Solomon A., Fryre D. (2008) Young children's recognition of the intentionality of teaching. *Child Development*, 79(5), 1237–1256.

Zmyj N., Daum M.M., Aschersleben G. (2009) The development of rational imitation in 9- and 12-onth-ld infant. *Infancy*, 14(1), 131-141.

Zmyj N., Buttelmann D., Carpenter M., Daum M.M. (2010) The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, 106, 208–220.

Zmyj N., Buttelmann D. (2014) An integrative model of rational imitation in infancy. *Infant Behavior and Development*, 37, 21-28.