RIVISTA INTERNAZIONALE DI FILOSOFIA E PSICOLOGIA
DOI: 10.4453/rifp.2015.0025
ISSN 2039-4667; E-ISSN 2239-2629
Vol. 6 (2015), n. 2, pp. 272-281

Studi

Implicit and Explicit Mentalization and its Relationship to Introspection

Giovanni Valeri

Ricevuto: 22 marzo 2015; accettato: 30 aprile 2015

Abstract In this paper I will discuss some claims made by Marraffa in his article *Mindreading and Introspection*. Early studies claimed that mentalization (or Theory of Mind – ToM) should first be observed during the preschool years. Subsequent research pointed out that ToM seems to be already present in infants. I will try to argue that the apparent inconsistency between these findings can be reduced by distinguishing between implicit and explicit ToM. From an evolutionary perspective the function of the two ToM systems seems to be different. The first is a genetically inherited neurocognitive mechanism which allows accurate expectations about behavior. The second is a culturally inherited skill, needed to modulate more complex social interactions. I will then discuss the relationship between implicit and explicit mentalization and introspection.
KEYWORDS: Theory of Mind; Implicit Mentalization; Explicit Mentalization; Introspection; Cultural Transmission.

Riassunto *La mentalizzazione implicita ed esplicita e la sua relazione con l'introspezione* – In questo articolo discuterò alcune tesi formulate da Marraffa nel suo *Mindreading and Introspection*. Studi precedenti hanno affermato che la mentalizzazione (o Teoria della Mente – TdM) dovrebbe essere osservata inizialmente durante gli anni precedenti la scolarizzazione. Ricerche successive hanno sottolineato che la TdM sembra essere già presente negli infanti. Cercherò di sostenere che l'apparente incongruenza tra questi risultati può essere ricomposta distinguendo tra una TdM implicita e una TdM esplicita. Da una prospettiva evoluzionistica la funzione dei due sistemi TdM sembra essere differente. Il primo è un meccanismo neuro cognitivo ereditario che permette di avere precise attese sul comportamento. Il secondo è un'abilità ereditata culturalmente, che ha bisogno di modellare più complesse interazioni sociali. Cercherò quindi di discutere il rapporto tra la mentalizzazione implicita ed esplicita e l'introspezione.
PAROLE CHIAVE: Teoria della mente; Mentalizzazione implicita; Mentalizzazione esplicita; Introspezione; Trasmissione culturale.

Introduction

OUR ABILITY TO ASCRIBE MENTAL states to ourselves and others is known as "mentalizing", "theory of mind", "folk psychology" or "mind reading." It has been a major focus of

G. Valeri - Dipartimento di Neuroscienze, IRCSS Ospedale Pediatrico Bambino Gesù, Piazza Sant'Onofrio, 4 - 00165 Roma (I)

E-mail: giovanni.valeri@opbg.net (✉)

philosophical investigation for centuries and of scientific enquiry for over 30 years.

The lack of consensus on how to characterize this human capacity to reason about mental states (such as belief) has been highlighted by recent research. Children under 3 or 4 years of age fail critical tests of belief reasoning (explicit ToM), yet infants apparently pass implicit false belief tasks at 13 or 15 months and infants as young as 7 months seem to be capable of mind reading. Non-human animals also fail critical tests of belief reasoning but may demonstrate very complex social behaviours.[1] Studies indicating that infants are capable of mindreading apparently support the view that mind reading depends on genetically evolved mechanisms since there is very little opportunity for cultural inheritance in the first months of life.

However, other research suggests that, although infants seem to be mind reading, they are not using the same mechanisms that control "full-blown" or "explicit" mind reading in adults – mechanisms that allow us to deliberate and talk about mental states.[2]

Research on infants, which infer evidence for mind reading from nonverbal behavior such as looking time and anticipatory looking, are said to provide evidence for *"implicit" mentalization***.** Implicit mentalization can be interpreted in TWO ways: the *continuity interpretation* (One-ToM system) or the Two-ToM systems hypothesis.

(1) The *continuity interpretation* (One ToM system) suggests that implicit mentalization is controlled by the same neurocognitive mechanisms that mediate explicit mind reading in adults.[3] Scholars of the One-ToM Hypothesis (One-ToM) propose that human beings operate with one and only one mentalizing system – which is understood as a cognitive architecture of a particular design and a dedicated, domain-specific function. Carruthers has advanced the most developed contemporary version of the One-ToM hypothesis, which holds that one, and only one, mind reading system operates throughout the whole of human development – from early infancy to adulthood. Crucially, as Carruthers emphasises, «while the operations of this system probably become more streamlined and efficient with age, its representational capacities do not alter in any fundamental way».[4]

(2) The Two-ToM systems interpretation suggests that implicit and explicit mind reading arise from different neurocognitive mechanisms; both systems are domain specific, specialized for thinking about mental states.[5] The Two-ToM hypothesis proposes that humans, at least, may be operating not with one, but with two functionally distinct mindreading systems. One of the main motivations behind this hypothesis is apparently the fact that mastery of propositional attitude concepts and their attribution require a great deal of cognitive sophistication. Two-ToMH postulates that in normally developing human adults implicit minimal ToM and explicit full ToM continue to exist intact and operate alongside one another.

The Two-ToM hypothesis has adequate means to deal with the developmental "paradox". On the assumption that implicit minimal ToM comes into play early on, it is hypothesized that human infants use this system, and only this one, when attributing mental states in cases of basic social cognition. Another mentalizing capacity comes into play only after older children begin to pass explicit, verbally based false belief tests – a first sign of the emergence of full explicit ToM.

Another reason to believe in the Two-ToM hypothesis, stems from the need to explain a range of evidence about human adult performance which suggests that ToM responding is sometimes fast and automatic and at other times slow and effortful. Moreover, it seems that the operation of explicit Full ToM abilities in adult humans is sometimes affected by a more automatic tendency to engage in basic ToM tasks in certain experimental set ups.

The Two-ToM hypothesis also has the advantage of being able to explain why implicit basic ToM abilities are widespread, oc-

curring not only in human adults under cognitive load but also in infants and in other non-human animals, whereas explicit full ToM cognition is comparatively rare.

According to the two ToM hypothesis, the implicit system develops early and tracks mental states in a fast and efficient way, whereas the explicit system develops later, operates more slowly, and makes heavier demands on executive functions, such as working memory and inhibitory control. Evidence for dissociation between implicit and explicit mind reading is found in studies of neurotypical adults.

In tasks in which adults make verbal judgments about others people's thoughts and feelings (explicit mind reading), judgment accuracy is impaired by concurrent performance of an executive function task. Instead, concurrent demands on executive function do not interfere with implicit mind reading. Further evidence for this dissociation is found in studies with autistic individuals, in which explicit mind reading can be achieved, in spite of continuing problems with implicit mind reading.

These dissociations are hard to reconcile with the continuity hypothesis (One ToM) but are compatible with the Two-ToM-systems interpretation. According to the Two-ToM systems account, moreover, explicit mind reading, which allows us to deliberate about mental states and to express our thoughts about mental states in words, develops slowly and is cognitively demanding. Indeed, specialization of mentalization continues into late adolescence and the performance on explicit tests of mind reading (including perspective taking, emotion recognition, and detection of pretense and irony) continues to improve between adolescence and adulthood.

In an evolutionary perspective, the function of the Two-ToM systems seems to be different. The first one is an expression of a basic motivation for social interaction: the infant is equipped with a *genetically inherited* neurocognitive mechanism that yields accurate expectations about behavior. The second one is a *culturally inherited* skill, needed to modulate more complex social interactions and for the transmission of culture.

Explicit mentalization

Evidence that the development of explicit mind reading depends on a slow process of learning, rather than on the maturation of genetically inherited neurocognitive mechanisms, comes from a twin study.[6] When more than one thousand twin pairs were given a comprehensive battery of explicit mind-reading tests at 5 years of age, the correlation in performance within pairs was the same for non-identical twins and for identical twins. This indicates a "substantial shared environmental influence but negligible genetic influence on individual differences in theory of mind". However, by itself, this twin study does not tell us about the nature of the environmental influence or about the kind of learning involved in the development of explicit mind reading.

Studies of social influences on the development of children's ToM ability can help us better understand the development of mentalizing, in the third-person and in the first-person, and the construction of the self representation. In the last decades, substantial research has contributed to a better understanding of social influences on the development of (explicit) ToM. Since the landmark study of Dunn and colleagues on the relationship of family environment to children's ToM ability, there has been increasing evidence suggesting that specific features of the early social environment are associated with precocity in children's understanding of mind. Dunn and colleagues reported a relationship between certain types of family interaction (such as the tendency to discuss feelings and use causal state language) and children's subsequent (explicit) ToM performance.

Subsequently, Perner and Ruffman reported that the mere presence of (older) sib-

lings had a facilitatory effect on (explicit) ToM performance. Indeed contact with older children and adults, beyond the nuclear family, appears to have a similar effect in aiding children's understanding of mind.[7]

Now it is important to deepen the relationship between research on social influences on (explicit) ToM and general theories about the development of ToM. There are two well-established accounts of mind reading: simulation theory and theory-theory.[8] Although they were previously treated as competing approaches to the explanation of mind reading, it is now widely accepted that these are complementary accounts, with *theory-theory* having a greater role in explicit high-level mind reading, whereas *simulation theory* is more relevant to implicit low-level mind reading.

This complementarity emerges also from neuroimaging studies: F. Van Overwalle, in a meta-analysis based on over 200 fMRI studies, founded that several brain areas process information relevant for social cognition, the capacity to understand people's behavioral intentions, beliefs, and personality traits.[9] The results suggest that inferring *temporary states* such as goals, intentions, and desires of other people – even when they are false and unjust from our own perspective – strongly engages the *temporo-parietal junction (TPJ)*. Inferring more enduring dispositions of others and the self, or interpersonal norms and scripts, engages the *medial prefrontal cortex (mPFC)*, although temporary states can also activate the mPFC.

Thus, the available evidence is consistent with the role of a TPJ-related mirror system for inferring temporary goals and intentions at a relatively perceptual level of representation, as described by the simulation model, and the role of the mPFC in a system that integrates social information across time and allows reflection and representation of traits and norms, and presumably also of intentionality, at a more abstract cognitive level, in accordance with the theory-theory account.

I think that the development of explicit ToM, in addition to the contributions of the theory-theory and simulation accounts, may be better understood according to a *socio-constructivist* model.

In a cultural evolutionary framework, the socio-constructivist view suggests that the novice's ideas about the mind are derived primarily not from simulation or from observation and hypothesis testing (according to theory theory) but from *instruction*, from what the novice is told about the mind by the expert mind readers in her social world. The social constructivist hypothesis is based on various findings, of which the most interesting are the development of ToM in deaf children and studies on cultural variations of ToM.[10]

Deaf children of hearing parents take 12 to 15 years to proceed through the same steps in the development of ToM that take hearing children 4 or 5 years. In contrast, deaf children of deaf parents, who learn sign language as their native language, do not show these delays. The deaf-of-hearing children have much less conversational experience than hearing or deaf-of-deaf children and this probably leads to the very delayed appearance of each developmental step, even though the sequence of these steps is the same.

In combination with other research involving typically and atypically developing children, research on deaf individuals and cross cultural studies indicates that we learn about the mind through both early socio-pragmatic interactions and through conversations about the mind. The appropriate conversational experience could come at first by listening to what expert mind readers say when they have no intention of teaching a novice. However, many studies suggest that experts, especially mothers, tailor or "epistemically engineer" their conversations about the mind so that it helps children to learn.

There are many data on the relationship between mindreading and language[11] but we should always remember that ToM development begins in infancy, before language development, as children begin to pay attention to other's minds following their eye-

gaze, engaging in joint attention and understanding other's goals and intentions. In summary, children culturally inherit from their parents, and other mindreading experts, such as siblings, the mechanisms specialized for the representation of mental states.

Debate currently surrounds whether mindreading is based on the preordained maturational unfolding of a neurobiological mindreading module or on the uniquely human early socio-pragmatic interactions and conversational experiences that all societies provide for their young to nurture children's developing understanding.

In contrast with nativist theories of mindreading (useful for understanding implicit ToM), the social constructivist, cultural evolutionary hypothesis suggests that humans do not genetically inherit neurocognitive mechanisms specialized for the development of explicit mind reading. Nonetheless, the genetically evolved mechanisms of implicit ToM provide much of the raw material, the genetic "start-up kit" for the construction of explicit mind reading: the mechanisms that become specialized for the representation of mental states and the processes that make cultural inheritance possible.

Implicit mindreading emerges from observing the behavior of others (and the context in which it takes place). This is a one-way process. The learner observes the actor, who need not be aware that he or she is being observed. In contrast, in a socio-constructivist, cultural evolutionary account, *explicit* mind reading emerges from the instructive behavior of others. This is essentially a two-way process. The behavior of the actor is designed to help the observer learn, and both actor and observer are actively engaged in a communicative process.

Components of a start-up kit for such processes might include the preference of newborn infants for faces, biological motion, eye contact and the preference of very young infants for objects that respond to their own actions with high contingency.

Explicit mindreading is culturally inherited: the neurocognitive mechanisms that allow us to deliberate and talk about mental states are probably constructed, or recycled, from mechanisms that evolved genetically to fulfill more general functions (e.g. to parse and predict dynamic sequences of events and to get information from others), and the construction process depends on *tuition*. Expert mind readers communicate mental-state concepts, and ways of representing these concepts, to novices. As the present generation of novices becomes expert, it passes on the knowledge and skill of mind reading to the next cultural generation.

From a Vygotskyian perspective most, possibly all, human neurocognitive skills are shaped by culture, and many are culturally inherited. Mindreading, implicit and explicit, is an essential aspect of human social intelligence; it evolved to provide an adaptive advantage in pursuing the aims of *two main motivational systems:* self-assertiveness / competition and cooperation. It is plausible to postulate a very strict link between the evolution of specifically human forms of explicit mentalization and the emergence of social systems that call for high cooperation.[12]

In infancy, when the enculturation process is just beginning, implicit mindreading mechanisms produce, under some circumstances, accurate expectations about the behavior of agents. Implicit mindreading mechanisms continue to operate throughout the life cycle, enabling swift social coordination of behavior when time is short and other demands on the neurocognitive system are heavy. It is possible that the outputs of these implicit mechanisms also contribute to the development of explicit mind reading by, for example, segmenting the stream of observable behavior into units that can subsequently be aligned with mental categories.

However implicit mindreading is radically insufficient for the development of explicit mind reading. Research suggests that no amount of *individual learning* – implicit mind reading, simulation, introspection, and watching the behavior of others – would be enough

for the development of explicit mind reading.

Introspection

These considerations are also relevant to the debate on the development of *introspection* discussed in Marraffa's article.[13] Many philosophers and psychologists, such as Aristotele, Agostino, Descartes, Locke, have traditionally assumed that *self knowledge* has a peculiar feature, that knowledge of our own thoughts (intentions, desires, opinions, beliefs) is direct and reliable.

Even today many philosophers argue that knowledge of at least a subset of our own thoughts, is based on direct access. This view is also widespread among cognitive scientists, in particular among those who believe that third-person mentalization is grounded in first-person mentalization. Other authors argue that self-knowledge comes from the act of turning on oneself the capacity to mindread other people. Let us consider, for example, Carruther's theory of introspection whih is center stage in Marraffa's paper.

In Carruthers' account, mentalization is one of the concept-using consumer systems in the global workspace models of human neurocognitive architecture proposed by Baars. Carruthers thinks that self knowledge is based on the observation of our own behavior and the context in which it takes place, and on the perception of our own emotional primary events (affects), other forms of sensory experience, visual images and inner monologue. Carruthers' theory is called "interpretive sensory-access" (ISA), a sophisticated version of the self/other parity account. Carruthers explains self-knowledge by direct sensory access and interpretation of own thoughts.

As Marraffa points out, there is a large amount of data supporting the ISA theory: in particular interpretations of the research on confabulation, cognitive dissonance and self-attribution in social psychology, data on "metacognition", data on neuroimaging, as well as criticism of "two methods" theories which predict, but have not confirmed, a dissociation between third-person mentalization and first-person mentalization in schizophrenia and autism.

In children with autism spectrum disorder, the capacity to attribute intentions to themselves is just as impaired as the capacity to attribute intentions to other people, and poor performances in both aspects result from the difficulties that such children have with mindreading in general. With regard to schizophrenia, passivity experiences are not best explained by the impairment of a system subserving first-person mindreading, but by a failure of the so-called "comparator system", one of the main components of the action-control system. If we accept the hypothesis that self-knowledge is based on turning on oneself the capacity to mindread other people, we should ask ourselves, referring to the previous discussion, if we are talking about implicit or explicit introspection (first-person mentalization).

We can assume an *implicit* system (unconscious) of *self experience* which includes, primarily, the basic emotional events (affects) felt by the subject and other forms of sensory experiences, basic aspects of motor intentionality, in addition to perceptions of visual images, and later, with the development of language, also interior monologue. This implicit form of self experience could be discussed in reference to models of early development of the self, like those of the interoceptive, "sentient self" of Craig, the ecological self of Neisser, the core self of Damasio and the minimal self of Gallagher.[14]

While the self is a popular topic in both cognitive neuroscience and psychology, the term is often used to discuss multiple different phenomena, and can thus be difficult to define.[15] Some of the most prominent and influential thinkers in psychology have theorized about the self. James wrote in *The Principles of Psychology* that the self is not a single primordial entity. This early conceptualization set the stage for later work examining multiple facets of the self.

Neisser claims that people have access to five different kinds of information about themselves. He describes five kinds of *self-knowledge* which may develop during different periods: (1) the ecological self, perceived with respect to the physical environment; (2) the interpersonal self, which depends on emotional and other species-specific forms of communication; (3) the temporally extended self based on memory and anticipation, implying a representation of self; (4) the private self, reflecting knowledge that our conscious experiences are exclusively our own; and (5) the conceptual self, based on sociocultural experience. According to this view, the self is not some special part of a person or brain, but rather a whole person considered from a particular point of view.

Gallagher delineates yet another distinction which he calls the "*minimal self*" versus the "*narrative self*". Here, the "*minimal*" self is referred to as the self devoid of temporal extension; phenomenologically, a consciousness of oneself as an immediate subject of experience, depending on brain processes and an ecologically embedded body. The "*narrative*" self, on the other hand, involves personal identity and continuity across time; it is a self-image constituted with a past and future in stories that we and others tell about ourselves.

Jeannerod's account is also interesting: he espouses the view that a key component of *self-recognition* in humans is recognizing oneself as the owner of a body and the agent of actions. These sensations of *ownership* and *agency* arise from the congruence of proprioceptive feedback and sensory signals from body parts, and central signals that contribute to the generation of movements. He claims that the sense of agency provides a way for the self to build an identity independent of the external world.[16]

These raw elements – *affective, motor, sensory, and quasi-sensory states such as visual imagery or inner speech tokens* – along with descriptions of themselves that children receive in their interactions, primarily with the caregiver, and then with "significant others" (siblings, peers, other significant adults), would constitute the basis for *self-knowledge* and, later, for *self-awareness*. The reference to sensorial data which the subject would access directly, according to the ISA model, would also show the importance of the fundamental representation of a *physical, bodily self*, to which a *social self* (what we are for and in relation to others), and finally, but only in certain socio-cultural contexts a psychic self, a self representation in the virtual space of the mind would be added, gradually, according to the William James' model.

The hypothesis of a *dissociation* between *bodily, social and psychological aspects of self-consciousness* is congruent with data from cultural psychology and ethnopsychiatry, which show the predominance of physical and social rather than psychological self-consciousness in adults belonging to preliterate cultures. The early evidence concerning the preliterate subjects' difficulties in representing a psychic self, an inner experiential space, is reported in A. Luria.[17]

Data from developmental psychology and neuroimaging studies are also congruent with the hypothesis of a *dissociation between bodily, social and psychological aspects of self-consciousness.* Gillihan and Farah suggest that physical or embodied self-related processes and psychological or evaluative self-related processes rely on distinct large-scale brain networks.[18] From a Vygotskian perspective, the development of (explicit) introspective self-consciousness is an «outward-in construction that occurs in an interpersonal context, namely, in the relationship with caregivers and peers».[19]

Introspection, understood in this theoretical context, is relevant to the psychodynamic topic of defenses, according to the hypothesis that our activity of re-appropriation of the products of the neurocomputational unconscious is governed by a self-apologetic defensiveness.[20] For example, in *Strangers to Ourselves* T.D. Wilson argues that the self-transparency assumption could make it easier for subjects to engage in various kind of adaptive self-deception, helping them to build

and maintain a positive self-image.[21]

In *Mindreading and Introspection* Marraffa argues that subjective identity develops through the act of turning on oneself the capacity to mindread other people through socio-communicative interaction with caregivers (and later other social partners) investigated by the psychodynamics theory: «The young child who turned his mindreading abilities upon himself under the thrust of the caregiver's mind-minded talk, by the end of pre-school years begins to grasp his introspective self-description as rationalized in terms of autobiography».[22]

This process of narrative self-construction includes a crucial psychodynamic aspect: affective growth and the construction of identity are inextricably linked. Thus, self-consciousness is a complex neuro-cognitive and psychosocial construction, which develops from the implicit, automatic and prereflective processing of representations of objects (*object-consciousness*), through *awareness* and then *self-awareness* of the body, up to explicit *introspective self-awareness* and then *narrative identity*.

The primary form of self-consciousness is bodily self-consciousness, probably structured as a nonverbal and analogic own-body representation: «[B]ut very soon it begins to be mediated by the verbal exchange with the caregiver. In other words, in our species the chimpanzee-style, purely bodily self-awareness is almost immediately outstripped and encompassed by a form of descriptive self-consciousness that is strictly linked to linguistic tools and social cognition mechanisms».[23]

Bodily self-description is the prerequisite for *autobiographical* encoding and storage, acting as a fixed referent around which personally experienced event memories begin to be organized.[24] This hypothesis has a significant impact on the psychopathological understanding of severe mental disorders, such as autism and schizophrenia.

Humans are pre-wired for interpersonal relationships from their birth, and implicit mindreading is part of such pre-organization. Introspective self-consciousness takes shape in the child in relationship with caregivers, in a form of scaffolding which initially is not linguistic, which is made up of non verbal interactions, turn taking, exchange of communicative gestures, affect regulation, basic socio-pragmatic communicative competence, and not only of words, descriptions, designations, evaluations of the person.

Through dialogue (nonverbal and verbal) with caregivers, and then with other social partners, children construct their own identity, both objective (for others) and subjective (for themselves):[25] «The intentional actions and attitudes repeatedly expressed towards the young child by caregivers and peers serve as the inferential basis for attributing generalized intentional properties to the self in an attempt to rationalize the social partner's self-directed behaviors».[26]

We can thus assume the existence of two different mechanisms: one at the base of the *implicit self experience* and one at the base of the *explicit self representation*. In normally developing human adults implicit self experience and explicit self representation continue to operate alongside one another, while in severe mental disorders we can find several kinds of impairments. For example, abnormalities in the awareness of action, and consequently impairments of the sense of agency and ownership in schizophrenia, or the impairment of *psychological* self awareness, self reference and representation of intentions in autism.

Explicit mentalization, in the third-person (mindreading) and in the first-person (introspection), requires a social constructivist analysis, as well as a neurocognitive one. This socio-constuctivist perspective on explicit first-person mentalization, accounts for various research findings. For example, studies by researchers working in the framework of the theory of "self-perception", initiated by Bem, and influenced by the theories of Ryle and Skinner, and by symbolic interactionism: the identity-for-itself can be said to derive from the identity-for-others – we see ourselves, and define ourselves, essentially *internalizing* the way in which others see and define us.

Self-awareness, self-consciousness, would thus be the outcome of a complex process of self experience, self monitoring and self representation (physical, social and, finally, psychological) to which both implicit and explicit components contribute: implicit perception of bodily sensations, emotions, motor intentionality and explicit psychosocial self representation in relation to other social a gents. The integration between these different levels of self representation is the base of an sufficiently stable, although constantly precarious, "*feel existing*".[27]

Notes

[1] See S. BARON-COHEN, H. TAGER-FLUSBERG, M.V. LOMBARDO (eds.), *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, Oxford University Press, Oxford (UK) 2013, pp. v-x.

[2] See V. SOUTHGATE, *Early Manifestations of Mindreading*, in: S. BARON-COHEN, H. TAGER-FLUSBERG, M.V. LOMBARDO (eds.), *Understanding Other Minds*, cit., pp. 3-18.

[3] See R. BAILLARGEON, R. M. SCOTT, Z. HE, *False-belief Understanding in Infants*, in: «Trends in Cognitive Sciences», vol. XIV, n. 3, 2010, pp. 110-118; P. CARRUTHERS, *Mindreading in Infancy*, in: «Mind and Language», vol. XVIII, n. 2, 2013, pp. 141-172.

[4] P. CARRUTHERS, *Mindreading in Infancy*, cit., p. 142.

[5] Cfr. I.A. APPERLY, *Mindreaders: The Cognitive Basis of "Theory of Mind"*, Psychology Press, Hove and New York, 2011; S. BUTTERFILL, I.A. APPERLY, *How to Construct a Minimal Theory of Mind,* in: «Mind and Language», vol. XXVIII, n. 5, 2013, pp. 606-637; J. PERNER, *Who Took the Cog out of Cognitive Science? Mentalism in an Era of Anti-cognitivism*, in: P.A. FRENSCH, R. SCHWARZER, (eds.), *Cognition and Neuropsychology: International Perspectives on Psychological Science*, Vol. I, Psychology Press, London 2010, pp. 241-261; C.M. HEYES, C. FRITH, *The Cultural Evolution of Mind Reading,*, in: «Science», vol. CCXLIV, n. 6190, 2014, pp. 12430911-12430916.

[6] See C. HUGHES, S. R. JAFFEE, F. HAPPÉ, A. TAYLOR, A. CASPI, T. MOFFITT, *Origins of Individual Differences in Theory of Mind: From Nature to Nurture?*, in: «Child Development», vol. LXXVI, n. 2, 2005, pp. 356-370.

[7] See J. DUNN, J. BROWN, C. SLOMKOWSKI, C. TESLA, C., L. YOUNGBLADE, *Young Children's Understanding of Other People's Feelings and Beliefs: Individual Differences and Their Antecedents*, in: «Child Development», vol. LXII, n. 6, 1991, pp. 1352-1366; J DUNN, M. BROPHY, *Communication, Relationships and Individual Differences in Children's Understanding of Mind,* in: J.W. ASTINGTON, J. BAIRD (Eds.), *Why Language Matters for Theory of Mind*, Oxford University Press, Oxford (UK) 2005, pp. 50-69; J. PERNER, T. RUFFMAN, S. LEEKAM, *Theory of Mind is Contagious: You Catch it From your Sibs*, in: «Child Development», vol. LXV, n. 4, 1994, pp. 1228-1238; J. CARPENDALE, C. LEWIS, *Constructing an Understanding of Mind: The Development of Children's Social Understanding within Social Interaction*, in: «Behavioral and Brain Sciences», vol. XXVII, n.1, 2004, pp. 79-96.

[8] See A. GOPNIK, H.M. WELLMAN, *Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory,* in: «Psychology Bulletin», vol. CXXXVIII, n. 6, 2012, pp. 1085-1108; A.I. GOLDMAN, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* Oxford University Press, Oxford 2006.

[9] See F. VAN OVERWALLE, *Social Cognition and the Brain: A Meta-analysis*, in: «Human Brain Mapping», vol. XXX, n. 3, 2009, pp. 829-858.

[10] See H. WELLMAN, C. PETERSON, *Theory of Mind, Development, and Deafness*, in: S. BARON-COHEN, H. TAGER-FLUSBERG, M.V. LOMBARDO (eds.), *Understanding Other Minds*, cit., pp. 51-71; H.M. WELLMAN, F. FANG, C. PETERSON, *Sequential Progression in a Theory-of-mind Scale: Longitudinal Perspectives*, in: «Child Development», vol. LXXXII, n. 3, 2011, pp. 780-792; A. SHAHAEIAN, C. C. PETERSON, V. SLAUGHTER, H. M. WELLMAN, *Culture and the Sequence of Steps in Theory of Mind Development*, in: «Developmental Psychology», vol. XLVII, n. 5, 2011, pp. 1239-1247.

[11] See J.W. ASTINGTON, J. BAIRD (eds.), *Why Language Matters for Theory of Mind*, cit.

[12] See R. BOYD, P. J. RICHERSON, J. HENRICH, *The Cultural Niche: Why Social Learning is Essential for Human Adaptation*, in: «*Proceedings of the National Academy of Science - U.S.A.*», vol. CVIII (suppl. 2), 2011, pp. 10918-10925; P.J. RICHERSON, R. BOYD, *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago 2008; M. TOMASELLO, *A Natural History of Human Thinking*, Harvard University Press, Cambridge (MA), 2014.

[13] See P. CARRUTHERS, *Mindreading the Self*, in: S. BARON-COHEN, H. TAGER-FLUSBERG, M.V. LOMBARDO (eds.), *Understanding Other Minds*, cit., pp. 467-485; E. SCHWITZGEBEL, *Introspection, What?*, in: D. SMITHIES, D. STOLJAR (eds.), *Introspection and Consciousness*, Oxford Scholarship Online, Oxford 2012, pp. 29-48 – doi: 10.1093/acprof: oso/9780199744794.003.0001.

[14] See See A. CRAIG, *How do you Feel? Interoception: The Sense of the Physiological Condition of the Body*, in: «Nature Reviews Neuroscience», vol. III, 2002, pp. 655-666; A. CRAIG, *The Sentient Self*, in: «*Brain Structure and Function*», vol. CCXIV, n. 5-6, 2010, pp. 1-15; U. NEISSER, *Five Kinds of Self-knowledge*, in: «*Philosophical Psychology*», vol. I, n. 1, 1988, pp. 35-59; U. NEISSER, *Criterion for an Ecological Self*, in: P. ROCHAT (ed.) *The Self in Infancy: Theory and Research*, Elsevier, Amsterdam 1995, p. 17-34; S. GALLAGHER, *Philosophical Conceptions of the Self: Implications for Cognitive Science*, in: «Trends in Cognitive Sciences», vol. IV, n. 1, 2000, pp. 14-21; A. DAMASIO, *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*, Random House, New York 2000.

[15] See G. JERVIS, *Fondamenti di psicologia dinamica*, Feltrinelli, Milano 1993; G. JERVIS, *Presenza e identità,* Garzanti, Milano 1984.

[16] See M. JEANNEROD, *The Mechanism of Self-recognition in Humans*, in: «Behavioural Brain Research», vol. CXLII, n. 1, 2003, pp. 1-15.

[17] See A. LURIA**,** *Cognitive Development: Its Cultural and Social Foundations* (1974), Harvard University Press, Cambridge (MA) 1976; A. LILLARD, *Ethnopsychologies: Cultural Variations in Theories of Mind*, in: «Psychology Bulletin», vol. CXXIII, n. 1, 1998, pp. 3-32.

[18] See S. GILLIHAN, M. FARAH, *Is Self Special? A Critical Review of Evidence from Experimental Psychology and Cognitive Neuroscience*, in: «Psychological Bulletin», vol. CXXXI, n. 1, 2005, pp. 76; M.D. LIEBERMAN, *Social Cognitive Neuroscience: A Review of Core Processes*, in: «Annual Review Psychology», vol. LVIII, 2007, pp. 259-289; J.H. PFEIFER, S.J. PEAKE, *Self-development: Integrating Cognitive, Socioemotional, and Neuroimaging Perspectives*, in: «Developmental Cognitive Neuroscience», vol. II, n. 1, 2012, pp. 55-69.

[19] M. MARRAFFA, *Mindreading and Introspection*, in: «Rivista Internazionale di Filosofia e Psicologia», vol. VI, n. 2, 2015, pp. 249-260, here p. 255.

[20] See G. JERVIS, *The Unconscious*, in: M. MARRAFFA, M. DE CARO, F. FERRETTI (eds.), *Cartographies of the Mind. Philosophy and Psychology in Intersection*, Springer, Berlin 2007, pp. 147-158, here p. 150; M. MARRAFFA, A. PATERNOSTER, *Sentirsi Esistere. Inconscio, coscienza, autocoscienza.* Laterza, Bari-Roma 2013.

[21] See T.D. WILSON, *Strangers to Ourselves*, Harvard University Press, Cambridge (MA) 2002; D.M. WEGNER, *The Illusion of Conscious Will*, MIT Press, Cambridge (MA) 2002.

[22] M. MARRAFFA, *Mindreading and Introspection*, cit., p. 257.

[23] M. MARRAFFA, *The Unconscious, Self-consciousness, and Responsibility*, in: «Rivista Internazionale di Filosofia e Psicologia», vol. V, n. 2, 2014, pp. 207-220, here p. 212.

[24] See M. MARRAFFA, A. PATERNOSTER, *Disentangling the Self. An Outline of a General Theory of Self-consciousness*, in: «New Ideas in Psychology», in press.

[25] See M. MARRAFFA, *The Unconscious, Self-consciousness, and Responsibility*, cit., p. 216; M. MARRAFFA, C. MEINI, *L'asimmetria fra la prima e la terza persona: implicazioni per la teoria dell'attaccamento*, in: «Attaccamento e Sistemi Complessi», vol. II, n. 1, 2015, pp. 45-64.

[26] G. GERGELY, *The Development of Understanding Self and Agency*, in: U. GOSWAMI (ed.), *Blackwell Handbook of Childhood Cognitive Development*, Blackwell, Oxford 2002, pp. 26-46, here p. 42.

[27] See G. VALERI, R. WILLIAMS, *Inconscio e formazione dell'identità nei pazienti psicotici e autistici*, in: «Sistemi intelligenti», vol. XXVI, n. 1, 2014, pp. 103-118.