A Properly Embodied Self within a Naturalistic, Bottom-up and Systemic-Relational Framework

Massimo Marraffa †
massimo.marraffa@uniroma3.it

Tiziana Vistarini †
tvista3@gmail.com

ABSTRACT

In this article a neo-Jamesian approach to the self is developed within a naturalistic, bottom-up, and systemic-relational framework. In this approach, consciousness of the body as one's own body is a necessary precondition of self-consciousness as psychological self-awareness, and hence of a socially and historically situated narrative self. Thus we take on board the criticism of those accounts of the narrative self that pay little attention to embodiment, or go to the extreme of stating that the narrative self is abstract and hence not embodied. But at the same time, we reject the idea that the embodiment of the narrative self is provided by a *pre-reflective self-consciousness*. By contrast, we view self-consciousness as *a construction all the way down*, which develops from automatic and pre-reflective processing of representations of objects (object-consciousness), through *awareness* and then *self-awareness* of the body, up to introspective self-awareness and then narrative identity. This form of constructivism is a naturalistic reinterpretation of the Hegelian idea that selfhood is socially constructed and self-experience intersubjectively mediated. However, it enables to move away from that blend of narrativism and epiphenomenism that characterizes some strands of the socio-constructivist approach to the narrative self. The storied Me that the selfing process makes is not "empty chatter", but rather a causally efficacious layer of personality viewed as a self-unifying system.

1. Introduction

A number of books and articles on the development of the self have been published in recent philosophy of mind, witnessing the increasing interest in investigating the phenomena surrounding self-consciousness and the sense of

† University of Roma Tre, Rome, Italy.

self from a standpoint that integrates philosophical considerations with the relevant data in cognitive science (especially in developmental psychology). Bermúdez (1998) can be considered as a pioneer of this approach with its attempt to create a dialogue between analytic Kantianism (a line of thought that runs from Peter Strawson to Gareth Evans and Quassim Cassam) and cognitive science. In the same integrative vein, Musholt (2015) proposed an important model for the gradual transition from self-related information implicit in the nonconceptual content of perception and other forms of experience to the explicit representation of the self in conceptual thought.

Some aspects of these important works on self-consciousness and the sense of self are against the background of this article. The way they deal with the topic, however, is different from ours: in the end they are still works of analytic philosophy of mind, whereas our article will contain very little that is recognizable as conceptual analysis; rather, it can be viewed as an exercise in "theoretical psychology", along the lines of Carruthers' (2011) theory of self-knowledge.

Now, if we turn our attention to current psychological sciences, we immediately realize that today, more than a century away William James's ground-breaking chapter *The Consciousness of Self* (1890), "one cannot make much progress through most areas of human psychology without encountering constructs that invoke the self" (Leary & Tangney, 2012, p. vii). In the past 60 years hundreds of thousands of scholarly articles and chapters have been published about the self; in this article we focus on three factors that contribute to explaining why the topic of identity has played such a pivotal role in psychology (cf. Jervis 2011).

The first factor concerns general psychology, and consists in the inextricable link between identity self-description and self-consciousness. The second pertains to dynamic psychology and developmental psychology, and consists in the fact that the construction of affectional life, in the course of infancy and, subsequently, throughout one's entire life, is closely linked to the construction of an identity that is well-defined and accepted as valid. The third concerns social psychology, and consists in the fact that we constantly negotiate the validity of our identity in exchanges with other people. In this article we tap

these literatures to develop a neo-Jamesian approach to the self within a psychodynamic, socio-cognitive and developmental framework.[1]

The theoretical backbone of this approach is a specific interpretation of William James's classic theory of the self as constituted by the couple <I, Me>. Following Dan McAdams' insightful commentary, James is interpreted as arguing that the self-as-I (the self as a subject) is not "the inner psychological entity that is the center or subject of a person's experience" (Leary & Tangney, 2012, p. 5). The I is rather a *process*; "I" does not refer to a noun but to a *verb*: "it might be called 'selfing' or 'I-ing', the fundamental process of making a self out of experience" (McAdams, 1996, p. 302). By contrast, James' self-as-Me (the self as object) is "the primary product of the selfing process"; it is "the self that selfing makes" (*ibid.*). The Me exists as an evolving collection of self-attributions (James' material, social and psychological selves) which originate from the I-ing process. It is "the making of the Me that constitutes what the I fundamentally is" (McAdams & Cox, 2010, p. 162).

Thus, in contrast to those philosophical views that take self-consciousness as a basic modality of consciousness, a primary and simple "knowing of being-there"[2], James defines self-consciousness in terms of *identity*: it is a knowing of being-there *in a certain way*, a self-describing, an identity forming, which is "a unifying, integrative, synthesizing process" (McAdams, 1997, p. 56). So interpreted, James' theory of the duplex self anticipates a number of theories in developmental and personality psychology that have made appeal to a general organismic process for integrating subjective experience. Ryan (1995, cit. in McAdams, 1997) mentions Heinz Werner's orthogenetic principle, Jean Piaget's organization, and Carl Gustav Jung's individuation -- despite all their

[1] This essay adds to a series of papers on the self: Chiaradonna & Marraffa, 2018; Di Francesco, Marraffa, & Paternoster, 2016, 2019a,b; Guerini & Marraffa, 2017; Marraffa & Paternoster, 2016; Marraffa & Meini, 2018, 2019.

[2] The main reference here is to what Kant asserts in a famous passage of the first *Critique*. As is well known, Kant agrees with Hume that the empirical apperception "can give us no constant or enduring self in the flow of inner appearances" (Kant, 1998, p. 232, A 107). Yet, he thinks that one may shift from the analysis level of psychological experience to that of transcendental arguing, and here posits a pure apperception: "I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am", he writes in the first Critique (B157); and in B158 he adds that "[t]he consciousness of self is [...] far from being a knowledge of the self" -- that is, the consciousness of existing is distinguished from the consciousness of existing in a certain way. Thus Kant's I think ("that accompanies all my representations") is something undetermined and void ("a something=X"), which, not unlike Descartes' cogito, lays a claim to being a *primum*.

differences, these constructs share the idea that human experience tends toward "a fundamental sense of unity in that human beings apprehend experience through an integrative selfing process" (McAdams, 1997, p. 57).

The integrative selfing process gives rise to different kinds of unity, corresponding to different and co-constituting forms of human selfhood. The most minimal form of the Me is *bodily* self-consciousness, which consists in the capacity to construct a representation of one's own body as an entire object, simultaneously taking this representation as a subject, i.e., as an active source of the representation of itself. At a more advanced level of complexity, the Me is *psychological* self-consciousness, which is the introspective recognition of the presence of the virtual inner space of the mind, separated from the other two primary experiential spaces, i.e. the corporeal and extracorporeal spaces. Psychological self-consciousness will evolve into the most cognitively demanding form of self: a narrative self.

2. Preliminaries: the construction of the self within a naturalistic, bottom-up, and systemic-relational framework

Our neo-Jamesian account of the self is developed within a *naturalistic*, *bottom-up*, and *systemic-relational* framework.

The theory is *naturalistic* in the Quineian sense that it professes "a resolute skepticism in the face of any 'higher level' of inquiry that purports to stand above the level of ordinary science" (Maddy, 2001, p. 39). If in the Kantian scheme there are the methods of science, at the empirical level, and the methods of transcendental analysis, at the transcendental level, the naturalistic philosopher sees herself as a member of the scientific community; she regards the methods and techniques of science as the best way to find out about the world. As Quine puts it, "it is within science itself, and not in some prior philosophy, that reality is to be identified and described" (Quine, 1981, p. 21). In the light of this rejection of Kant's two-level scheme, the synthesizing selfing process should be viewed as the activity of a *psychobiological* system, and not as a Kantian synthetic function. In Kant's a priori philosophical psychology, the person is always given in its unity, as if the psychological level of analysis were always and in any case guaranteed by the transcendental level of analysis. This, however, does not hold true in light of the empirical theorizing of science: in the case of the synthesizing selfing process, it will be argued, the *empirical* subject is primarily *non-unitary*,

and gains a sense of unity in the act of mobilizing resources against the threat of disintegration.

Naturalism brings us to the idea of a *bottom-up* methodology. A developmental approach to the self is pursued, which attempts to reconstruct how the complex psychological functions underlying the adult self-conscious mind evolve from more basic ones. This approach does not appeal to our introspective self-knowledge, but rather to the results of investigations into the gradual construction of human self-awareness. In this perspective, the study of the 0-1 year-old infant's subjectivity should follow the example of the study of animal subjectivity, where cogent evidence can be found of very complex inter-individual behavioral dynamics that are produced by conscious (but not self-conscious) activities of representation. Animal behavior researchers (and especially primatologists) "are typically circumspect in their interpretations, limiting their claims to operationalizable terms [...] rather than making claims about the nature of the experience that may be involved in an animal's performing a task" (Allen & Trestman, 2017, §7.4). Recently, cognitive neuroscience has shown how to investigate the 0-1 year-old infant's subjectivity limiting one's claims to operationalizable terms. The groundbreaking study by Kouider et al. (2013) showed that one-year-old children have a brain signature similar to that associated with conscious perception in adults, albeit much slower (reflecting the delayed myelination and immaturity of the young brain)[3].

Our framework, however, aims to avoid not only a top-down ontologically inflationary approach to the self, but also an overly reductionist approach which explains *everything* in terms of bottom-up neurocognitive mechanisms. This is where a *contextualist* and *systemic* perspective comes into play. Here the individual's psychological problems are investigated by putting them in the inter-individual and social context in which they arise and obtain a sense. This systemic naturalism is rooted in the Chicago school of functionalism, and – as we will see immediately – is the foundation of attachment theory, namely, the psychodynamic tradition within which our neo-Jamesian account of the self is built.

[3] One of these brain signatures of conscious perception is the P300 slow wave, a particular type of electric wave that occurs whenever an adult subject is attending to a consciously perceived picture or sound. These signals start roughly around 300 milliseconds after the onset of the image or sound, can be long-lasting, are depolarizing (positive) relative to a reference electrode, and are particularly prominent above the frontal lobe.

2.1. Motivation and attachment

A distinguishing mark of the development of post-Freudian psychoanalysis is the focus on relational themes, especially on the forms of cognitive-affectional relationality of the very young child. The rise of attachment theory is part of this orientation. This theory hinges on two psychological constructs, motivation and attachment, which have played a central role in fostering an exchange between psychoanalysis and psychological sciences. Such a role must be viewed within the context of a deep revision of the anthropology underlying Freud's psychoanalysis. According to this traditional (Hobbesian) conception of human nature, individuality exists prior to relationality; sociality is a reality that comes "after" individuality, since it is a cultural product generated by the necessity to live together.

During the last decades, however, biology, sociology and behavioral economics have productively interacted with psychological sciences, making it increasingly clear that human sociality is not something that originates only from culture, but is rather a dimension that belongs to the definition of the human individual itself. According to this new anthropology, human sociality complies with certain natural predispositions; individuals are seen as bearers of a very complex suite of *motivations*, which are always, and have been from the beginning, *relational*. According to Lichtenberg's (1989) well-known taxonomy, all these motivations give place to complex interactions between five "motivational systems": the need to fulfill physiological requirements; the need for attachment and affiliation; the need for assertion and exploration; the need to react aversively through antagonism and/or withdrawal; the need for sensual and sexual pleasure. It is to be noticed, however, that the aversive-aggressive system is largely dependent on the assertive-explorative one, whereas the sensual-sexual system depends largely on the attachment-affiliation one. Consequently, the fundamental motivational systems may be only two: one dedicated to self-assertiveness and competition, and another aimed to prosociality and cooperation (Jervis, 2001).

Lichtenberg's taxonomy of motivational systems delivers an anthropology that is neither pessimist nor optimist: human beings are naturally inclined to competition, and sometimes destructivity, but also to forms of sociality, cooperation and even altruism (e.g., Bowles & Gintis, 2011). Freud saw the precarious situations of compromise between social repression and drive discharge as conflictual and sources of uneasiness. By contrast, the spontaneous

situations of compromise that arise between the motivation to cooperate and the motivation to compete may turn out to be intelligent, well-organized and ingenious; and they are characterized not by uneasiness but by the generation of "non-zero-sum" relationships.

The claim of the primary nature of sociality is then the anthropological foundation of the psychodynamics of object relations and attachment. The primordial psychological need of the very young child, around which his mental life gradually takes shape, is not -- as Freud thought -- the oral drive gratification, but rather the physical contact and the construction of protective and communicative interpersonal structures.

Attachment is the primary matrix of cooperation, and in contemporary attachment theory and research, the dialectics between the attachment-affiliation system and the assertive-explorative system is the key to understanding the child's cognitive-affective development.[4]

2.2. Systemic naturalism

The overcoming of the traditional philosophical and psychological view of the human individual as an isolated primary subject, *a priori* "given" as autonomous, is the result of a contextualist and systemic perspective which puts the individual's psychological problems into the inter-individual and social context in which they arise and come to have sense. The theory of object relations seems to fully endorse this systemic approach to the study of relationality -- as Donald Winnicott puts it, what makes sense is not considering the infant in itself, but the *mother-infant dyad*. But a caveat is in order here.

With the adoption of a contextualist and systemic perspective, psychology draws inspiration from trends currently dominant in biology and sociology. There exists a long tradition in theoretical biology -- the so-called "developmental systems theory" (Griffiths & Tabery, 2013) -- in which the separation of the individual organism from the environment hardly makes sense. From this perspective, both the developments of Darwin's theory and the modern concepts of equilibrium, adaptation, innate/acquired interrelation and ecological niche are seen as leading us to consider the organism-environment

[4] At the center of attachment theory is the relationship between the "secure base" functions of the attachment figure and the individual's ability to explore the world and self, relatively free of anxiety. That is, one cannot comfortably engage in exploration (including self-exploration) without ties to other (i.e., a secure base). Cf. Holmes (2014); Eagle (2013).

structure as a single systemic whole, where neither of the two poles is primary with respect to the other, and thus, also to reject the contrast between the nature-based (i.e., inherited, genetically based) and nurture-based (i.e., acquired, environmentally mediated) characteristics in development. In animals as well as in human beings the development of the organism from the fertilized egg to reproduction and death consists in a series of structured interactions, each of which builds itself on the basis of the previous one, and each of which sees the interaction, on the one hand, of the onset of new environmental signals, and on the other, the gradual opening of new inner potentialities developed during the previous stages.

Now, in the case of the biological inspiration the consideration of psychological phenomena in terms of equilibria, and hence of systemic interactions, has a naturalistic origin and is in continuity with William James and the Chicago school of functionalism. Things change, however, when the systemic approach to the mind has a sociological origin. A forceful tendency has long existed in sociology and social psychology to attempt to make the investigation of human behavior more rigorously scientific by means of its *de-subjectivation* – hence the prevalent use of explanatory tools that have a structural-relational nature rather than dispositional-intentional one. We can already see this tendency at work in Talcott Parsons, with his turning Max Weber's typology of attitudes into a typology of role relations. Such role relations are always structured, and in dynamic equilibrium, and can hence be considered in an implicitly systemic perspective. Similarly, the evolution of areas at the interface of psychology and sociology like, since the 1960s, symbolic interactionism and the work of Erving Goffman, have resolutely pointed in the direction of a theory of the interactive construction of the description of the self and reality.

Now, what is primary in the systemic perspective is not the individual but the interaction, often viewed as communicative dynamic field. It may happen, then, that sociological inspiration makes such an approach more radical and ends up dissolving the individual. The result is a form of sociologism that neglects the value of systemic naturalism, that is, wipes out any sense of an ecological perspective in which the human organism is *biologically* part of the environment before being sociologically and culturally part of it. This antinaturalistic sociologism gives place to a "pure", disembodied relationalism, where the individual (the living and real information-processing organism) is

reduced to a mere knot in the tangle of an organized field of influences or, more properly, messages.

A good example of this unwelcome outcome is provided by those forms of sociolinguistic constructivism which completely dismiss psychological sciences, or seek to replace them with a "psychology of the surface" which is relational and linguistic, such that there are no information-processing mechanisms, not even mental states and processes: these things are opaque and unproductive; only relations and language hold. On this view, psychological phenomena are produced in social interaction, and above all in the context of "conversation", beyond which there is no mental process; mental processes are nothing but our conversational interactions. From here it is a short step to seeing persons not as the actors in or the agents of discourses, but rather as the products of the discursive practices themselves (e.g., Harré, 1986, 1987; and more recently, Carpendale & Lewis, 2006; Hutto, 2008). The self is thus entirely located within the public space of discourse.

Ironically, however, the suppression of the biological made by such antinaturalistic sociologism frustrates the very sense of integration that the systemic approach pursued, leading to the situation against which it aimed to struggle, that is, a conception in which (individual) biology and (social) relationality are split from each other, in that the former is deleted and the latter becomes all-encompassing.

Certainly, there is nothing in the theory of object relations that renders it ineluctably liable to such involution. Quite the contrary: it is wrong to think that if one speaks of the theory of object *relations*, then the theory is, as such, immediately *relational*. The idea of "object relation" is not strictly and *in itself* an *interactionist* theory, let alone a *systemic* theory. The subject can still be seen as *primary* with respect to the object. In other words, we can still have a relation in the traditional sense, namely in a one-directional sense; the theory of object relations is not necessarily a relational theory in the strict sense, i.e., a theory focused on the forms of an interactive dialectics that constantly generates new dynamic equilibria. That being the case, the different versions of the theory of object relations fit into different parts of the spectrum that from the classical conception of the subject seen as primary with respect to the object leads to the above rejected pure relationism. So we should not confuse and conflate the claims that minds are shaped by early interactions with others -- and that much that goes on in our mind has to do with our relationships with others and representations of these relationships (all claims that we can find in the theory of

attachment) -- with the radical, social-constructivist claim that "the basic unit of study" in psychoanalysis is not "the individual as a separate entity" but "an interactional field", which can be found in the relational theory of Stephen A. Mitchell (1988, p. 3).

2.2.1. Pure relationism in dynamicist style

Another path to pure relationism is a radical form of externalism that was put forward by some proponents of the dynamical approach to cognition (or "dynamicism").

According to some defenders of the dynamical approach to cognitive modeling, the dynamical analysis identifies the critical variables characterizing the state of a system and attempts to construct laws (a set of differential equations) to account for the system's trajectory through state space. The system can no longer be decomposed into subsystems (modules) that involve computations on representations. Consequently, the dynamical explanation is seen as incompatible with the explanatory style of the computationalist mechanism (cf. Chemero, 2009; and references in Chemero & Silberstein, 2008, pp. 11-13).

Most important for the current discussion, dynamicism dissolves the boundary between the cognitive system and the system's environment. Coupling between the equations describing a cognizing system and those describing the environment gives rise to complex "total system" behaviors. In this perspective, "the cognitive system is not just the encapsulated brain; rather, since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three" (van Gelder, 1995, p. 373). In this perspective, the role of the brain blurs in a conception of reality in which entities are undifferentiated variables and processes -- a Machian view, on some aspects (Marraffa, Paternoster, 2012, p.35).

In brief, dynamicism puts forward "the radical embodied cognition thesis": to understand the complex interplay of brain, body, and environment we do not need either the concepts of internal representation and computation or the mechanistic decomposition of a cognitive system into a multiplicity of inner neuronal or functional subsystems; all we need are the analytic tools and methods of dynamical systems theory (Clark, 1997, p. 148).

This revolutionary interpretation of dynamicism can be contrasted with Andy Clark's and William Bechtel's reformist projects, which aim to amend the computational-representational framework by drawing together insights from explanatory pluralism, mechanistic analysis, and dynamicism. Clark (1997, 2008) suggests that dynamical and computational-mechanistic explanatory patterns ought to interlock in a complete explanation of cognition, a claim that has been explored in depth by William Bechtel and his collaborators (e.g., Bechtel, 2008; Bechtel & Richardson, 2010; Kaplan & Bechtel, 2011; Bechtel & Abrahamsen, 2013).

In the early stage of the process of developing mechanistic models, scientists often assume that the processes that they are considering are performed serially. But when it is not possible for scientists to develop a linear model that is adequate to the phenomenon, they start to introduce feedback loops and other non-linearities in their attempts to develop adequate models. The outcome is what Bechtel and Richardson (2010) term "functionally integrated systems". As a result, a continuum emerges. At one end of the spectrum we have *fully decomposable* (or highly modular) systems, which are composed of subsystems that are completely independent except for the mutual exchange of outputs (this is the case with Fodor's encapsulated modules). If the interactions among the subsystems are weak but not negligible, the system is *nearly decomposable*. As the complexities of interaction among parts increase, the explanatory burden shifts from the parts (or, more precisely, the interactions within subsystems) to their organization (i.e., the interactions between subsystems). Thus we reach the other end of the spectrum, where we find *holistic* systems whose components are functionally equivalent and hence interchangeable. In between the nearly decomposable systems and the holistic ones, there are the integrated systems. In these systems, unlike the holistic systems, it is possible to isolate different parts that make distinctive contributions but also give rise to a complex set of interactions that are nonlinear, and hence much stronger than those of a nearly decomposable system.

Now, both Bechtel (2001) and Clark (1997) suggest that psychobiological cognition is likely to take up the intermediate space between nearly decomposability and holism, namely that of integrated systems; and in an integrated system, mechanistic analysis "provides the foundation for dynamical analysis" (Bechtel, 2001, p. 483) since the latter has explanatory force only insofar as it describes "the operations of the underlying mechanism" (Kaplan &

Bechtel, 2011, p. 443), only to the extent that it reveals "aspects of the causal structure of a mechanism" (Kaplan & Craver, 2011, p. 602). Bechtel and Abrahamsen (2010) refer to accounts integrating mechanistic decomposition of systems into parts and operations with the quantitative tools provided by dynamical systems theory as "dynamic mechanistic explanations".

This attempt to reconcile dynamical modelling and mechanistic analysis is particularly relevant here since Griffiths and Tabery (2013) have convincingly argued that the explanations at which developmental systems theory aims are mechanistic explanations, and often dynamical mechanistic explanations, of the developmental potential of the system.

3. The selfing process

With our naturalistic, bottom-up, and systemic-relational framework for conceptualizing the self in place, we can outline some milestones in the selfing process. First, the human neurocognitive system produces, over a period of about 15-18-24 months, a representation of the body as a whole – a "bodily self". Then an introspective experiential space is constructed, endowing the subject of a psychological self-consciousness. This is the result of the "affectivation" of bodily reflexivity first, and then of turning upon oneself a collection of other-directed social-cognitive abilities subserved by two early-developing neurocomputational systems: one underlying the psychological reasoning or "mindreading", the other underpinning sociomoral reasoning. Finally, with the development of autobiographical memory and autobiographical reasoning, psychological self-consciousness evolves in the ability to construct a self-narrative as a layer of personality.

3.1. Making the infant's bodily self-reflectivity an *affective* bodily self-reflectivity

The initial state in development is a condition of *undifferentiation* between self, other, and the world. From an observer's point of view, one sees the infant interacting with others and the world, but the infant still has to draw the distinctions between inner and outer, subject and object, and self and other. These distinctions will occur gradually, in the microsocial context of the interaction with the caregiver. This does not imply that the neonatal mind is devoid of subjectivity; it rather means that the newborn produces a rich

subjectivity, but, being immersed in it, cannot objectify it. She is an active subject in the sense of being a functional center organizing action, but she cannot "have" herself as an active subject. Her experiential space is purely *objectual*. Her affects, the need of contact, the oral yearning, the hunger, the possible gastric colic pains are not actually *hers* but are experienced as things and events in the same way as a light, a noise, a face above the cradle.

Once interpreted with the necessary methodological caution, the large amount of experimental data so often invoked to support the idea that infants under one year of age have "a primitive, proprioceptive form of self-consciousness already in place from birth" (Gallagher & Zahavi, 2019, §2) shows merely that they are access-conscious in the sense of being able to form first-order representations of objects and actions. In this perspective, when a baby of, say, six or eight months perceives, for example, her hand, she perceives it as an item in the objectual field, *not as a part of her body*. Indeed, in order to perceive it as a part of her body, she ought to possess the ability of representing her body as a whole, what is not the case: it is over the course of the first three years of life that "an explicit visuo-spatial representation of one's body progresses from *early awareness of individual body parts* to *representation of the body as a whole* in which the body parts together constitute a typical configuration that corresponds to others' bodies" (Brownell, Svetlova & Nichols, 2012, p. 40; emphasis added).

Clearly, the onset of a bodily form of self-consciousness requires not only that the infant becomes able to represent her entire body. Let us consider when the child -- between the ages of 15 and 24 months -- becomes able to recognize one's specular image in the mirror (see, e.g., Lewis & Brooks-Gunn, 1979; Nielsen, Dissanayake & Kashima, 2003). Mirror self-recognition involves being able to form a bodily image of oneself as an entire *object*, and simultaneously taking this image as a *subject*, i.e., as an active source of the representation of oneself. Here the subject recognizes a new kind of object of consciousness: the object is the subject itself, or better the objectified image of the subject -- "it is *me* there". That this marks the agent's achievement of self-objectivation as "me" is also supported by the evidence that mirror self-recognition is connected with the phenomena arguably associated with self-consciousness, such as verbal and deictic self-reference (Lewis & Carmody, 2008) or experiencing embarrassment, empathy and jealousy (Lewis, 2014). The gradual emergence of the capacity of mirror self-recognition during the second year of life indicates therefore the onset of a new modality of cognition

compared with the ability to build the image of any external object that is characteristic of animal consciousness in general.

Note that humans are not the only species that display mirror self-recognition; it has also been found in chimpanzees, orangutans, a few gorillas, elephants, dolphins, and magpies. In most cases, however, epistemological caution clearly suggests a *lean* interpretation according to which animals pass this test because of kinaesthetic-visual matching skills. Actually, such deflationary interpretation has been put forward also in the human case, as opposed to *rich* interpretations (e.g., children's mark-directed behavior is evidential of an introspective form of self-consciousness and a self-concept inherently linked to understanding the mental states of other people) and proposals lying somewhere between the two. Taking, as we do, mirror self-recognition as a marker of bodily self-consciousness falls within this last option (for references, see Kristen-Antonow et al., 2015, p. 2).

The acquired awareness of the body as one's own is the basic premise necessary to provide ourselves with that elementary reflexivity that allows us to know that we exist. Thus, self-consciousness in its most basic form, namely as awareness of one's own existence, is the cognizance of a physical identity. It rests not on a supposed pure and primary feeling of existing, but on a *self-describability* – the child gains access to the feeling of existing when she recognizes herself in a body distinguishable from others' bodies, when she comes to know herself as a bearer of physical, physiognomic, bodily features.

3.1.1. Making the infant's bodily self-reflectivity an *affective* bodily self-reflectivity

At an early stage this bodily self-consciousness is likely to be structured by a non-verbal, analogic representation of the physical self which is neither nonconceptual nor fully conceptual. However, in our species this "chimpanzee-style", purely bodily self-consciousness is soon outstripped and encompassed by a form of descriptive self-consciousness that is strictly linked to affective self-regulation. One hypothesis about the mechanisms underlying this transition is the social biofeedback theory of parental affect-mirroring (Gergely & Watson, 1996, 1999; Fonagy et al., 2002; Gergely, 2007; Gergely & Unoka, 2008; Gergely, Koós & Watson, 2010).

The model is completely at odds with the "strong intersubjectivist view", namely, the claim that infants are born with a pre-wired organization of their

minds that ensures a primary introspective access to their own affective and intentional mental states (Gergely, 2002). An example of such a position is Meltzoff and collaborators' hypothesis of a specific innate mechanism underlying intersubjective attributions during early imitative interactions. The affective behavioral acts of the other are mapped onto the infant's supramodal body scheme, allowing her to recognize the other person as "just-like-me" (Meltzoff, 2013). By imitating such acts, infants generate the corresponding feeling states in themselves; these are then *introspectively accessed* and attributed to the other by inference.

By contrast, the social biofeedback model begins with the hypothesis that at the beginning of life infants have a primary bias to attend to and explore the external reality, and construct representations mainly based on *exteroceptive* stimulation, while lacking a complementary capacity to cognize their internal world. With regard to that particular set of perceptual stimuli that are expressions of emotions, the model taps into the longstanding empirical research showing that young infants are already able to discriminate and respond to the caregiver's facial, bodily, and vocal displays of specific *basic emotions*.

In the Darwin-Tomkins-Ekman tradition, basic emotions (the most elemental among discrete emotions) are biologically based and pancultural packages of short-term, coordinated and automated responses to events in the environment, which include *somatic* components (e.g., measurable physiological changes), *motor* components (e.g., facial, bodily, and vocal expressions) and *motivational* components (i.e., action tendencies -- e.g., a flight tendency characterizing the fear response). These responses are assumed to be automatically elicited and coordinated by a causal, neurocomputational mechanism called the "affect program" which is elicited by automatic appraisals (e.g., Ekman, 1999). That young infants are innately able to recognize the overt manifestations of affect programs in other agents is what we mean when saying that they "detect" basic emotions -- or, in other words, that they are "sensitive to" affect programs.

Conversely, at this stage infants lack the *feeling* component of such discrete emotional states, i.e., there is no reflection of aspects of the other basic emotion components in infants' first-order consciousness. Initially, the set of internal visceral and proprioceptive cues that are activated when being in and expressing an emotion state 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). Rather, we can ascribe to very young infants

"the mental representation of bodily changes that are sometimes experienced as feelings of hedonic pleasure and displeasure with some degree of arousal" (Lindquist et al., 2012, p. 124). In other words, the original form of differentiation of the objectual experiential space is likely to occur in accordance with what is termed "core affect" (Russell, 2003; Barrett et al., 2007). In this perspective, information about the external world is translated into an internal affective code or state that functions as a kind of core knowledge about whether objects or events are helpful or harmful, rewarding or threatening, requiring approach or withdrawal. With awareness, core affect is experienced as feelings of pleasure or displeasure that are to some extent arousing or quieting.

In the transition from these primitive and simple feelings to the awareness of discrete emotional episodes -- e.g., "awareness of being "angry", rather than just experiencing some undifferentiated negative state of tension" (Gergely, 2007, p. 58) -- a fundamental role is played by protoconversational interactions (Bateson, 1979).

During protoconversations, well attuned adults are spontaneously inclined to mirror back the infant's affect-expressive displays in a "marked" way: in response to the child's realistic emotional expressions, adults tend to display rather congruent but schematic and often exaggerated behavior, frequently characterized by an incongruent element -- as, e.g., when the adult responds to a crying child with an expression that, while being initially very sad (as expected by a perfect mirroring), turns rapidly to a smile; or when he responds to a happy child with an expression mixing joy (i.e., the mirrored expression), tenderness and possibly surprise. And empirical data attest that the infant is able to register the high-but-not perfect degree of contingency[5] between the parental mirroring and her ongoing affective experience.

The expressive exaggeration of the parental mirroring, coupled with the soothing tone and the lack of the typical behavioral consequences of genuine expressions, fulfils a first function of mitigating what would have an excessive arousing effect for a young baby, still incapable of affective self-regulation. In particular, negative behavioral consequences for a sad infant are avoided or

[5] This process of "referential anchoring" depends indeed on the functioning of the so-called "Contingency Detector" (Gergely & Watson, 1996, 1999), an innate computational mechanism that would enable the infant to analyze the conditional probability of three contingent relations -- temporal contingency, spatial similarity and correspondence of relative intensity -- between own actions and effects in the external environment.

mitigated by a "non-fully sad response", thus escaping a vicious circle the infant would be unable to escape from.

At the same time – and most importantly in the present context – marked expressions have a crucial *pedagogical* function (Csibra & Gergely, 2011). By simultaneously making salient central aspects of the somatic emotional manifestations and signalling that the displayed emotion is "not for real", the adult encourages the child to "decouple" the emotional expression from its apparent referent. Once decoupled, however, the affect-mirroring display still needs to be interpreted by the infant as referring to "someone's emotion". The adult's gaze, ostensively and continuously directed to the infant, helps him to "*referentially anchor* the marked mirroring stimulus as expressing his *own* self-state" (Gergely & Watson, 1996, p. 1199). In such process, the child becomes progressively more and more sensitive to her emotional state, which can now recognize in its multiple, distinct components. The expression "social biofeedback" used to designate such process aims at recalling the terminology used in the physiological domain to denote what happens when, e.g., someone is sensitized to his arterial pressure through being exposed to a monitor displaying continuously the internal situation. In the social domain, the outcome of the same kind of process is the phenomenology of discrete emotions.

Thus, the parental affect-mirroring serves mainly two functions. A function of *sensitization*: the infant becomes sensitive to the set of internal physiological and proprioceptive cues that are active while her affect-expressive behavior is controlling the adult's marked affect-mirroring expressions. A function of *representation building*: the separate representations of the caregiver's affect-mirroring displays become associated with the infant's primary and procedural affective states; thus they form *secondary representations* that are about those primary affective states and provide the basis for the infant's emerging ability to control her emotion states. This is therefore a version of a higher-order theory of consciousness, where first-personal access to one's own mental life is made possible through socially mediated second-order representations.

To recapitulate. The earliest form of differentiation of the infant's objectual experiential space occurs in virtue of feelings of pleasure or displeasure that are to some extent arousing or quieting. In contrast, there is no phenomenology associated to basic emotions. In the initial stage, basic emotions are packages of somatic, motor and motivational components elicited and coordinated by causal mechanisms (affect programs) which play the role of social signals in the "negotiation" between infant and caregiver (Griffiths & Scarantino, 2009, p.

446). It is affect mirroring that adds a phenomenological component to basic emotion packages. As seen, marked mirroring displays are interpreted self-referentially by the infant, leading to their referential anchoring (in the form of internalized second-order representations) to those procedural basic emotion states that the mirroring displays contingently reflect. This process will lead to the internalization of discrete emotions into the infant's own inner life when – in the second year of life – the phenomenology of basic emotions is embedded into bodily self-consciousness, making the infant's bodily self-image an *affective* bodily self-image.

Two aspects of this socio-constructivist approach to affective introspection are particularly important for our purposes. First, it contributes to an anti-Cartesian agenda (Carruthers, 2019). All along the process leading to the construction of psychological self-knowledge, the child recruits some other-directed competence and redirects it to understand herself. Second, it supports the claim that bodily self-consciousness is a necessary premise of the ability to identify the presence of an inner experiential space. While not committing to the Jamesian idea that all emotions are perceptions of aroused states of the body (Damasio, 1999; Prinz, 2004), the earliest cognition of mental events appears to be the outcome of the acquired capacity to interpret "primary somatic data specific to categories of affective states and of attributing them to the self" (Hernik, Fearon & Fonagy, 2009).

3.2. Expanding the introspective self-consciousness

After internalizing basic emotions into one's own unfolding inner life, the child must learn to recognize and attribute to herself other kinds of mental states and activities, as well as forming the conceptual network that links such phenomena. It can be assumed that this expansion of the introspective experiential space takes place through the process of turning upon oneself a collection of other-directed social-cognitive abilities subserved by two early-developing neurocomputational systems: one underlying the psychological reasoning or "mindreading", the other underpinning sociomoral reasoning (Baillargeon et al., 2014; Buyukozer Dawkins et al., in press). In the mindreading and sociomoral domains, therefore, we find the same dissociation between other- and self-directed capacities that we found in the affective domain, where an other-directed affective competence scaffolds, in the protoconversational context, the development of a corresponding self-directed competence.

This hypothesis is fully consonant with attachment theory since, from the latter perspective, the mindreading and sociomoral reasoning systems may well be part and parcel of our being pre-wired to the interpersonal relationship from birth. It is to be noticed, however, that this mindreading system is a social-cognitive evolutionary adaptation that is independent of Bowlby's innate infant-caregiver attachment system. This is in contrast with the hypothesis, variously put forward by a number of attachment theorists and infant researchers, that there is a direct causal and functional link between early infant-parent secure attachment on the one hand, and the development of mindreading on the other (Gergely & Unoka, 2008).

According to Meins (2011, 2013), however, the observed link between security of the infant-parent attachment relationship and mindreading may be *indirect*, with both attachment security and mindreading performance being predicted by caregivers' *mind-mindedness*, i.e., the parent's proclivity in speech to attribute appropriate internal states to the child. In the first year of life, mind-mindedness is operationalized in terms of the appropriateness of parents' comments about their infants' internal states. This measure predicts various positive aspects of children's development, including in fact attachment security (e.g. Meins et al., 2012) and mindreading performance (e.g. Kirk et al., 2015).

Now, there is no doubt that caregiver-infant communicative interaction impacts on the development of mindreading; the problem is *how* the child's exposure to such interaction can have such an impact. In particular, the specific role that language plays in this context should be clarified. Jill and Peter de Villiers, for example, would disagree with Meins' hypothesis that language, in the form of appropriate mind-related comments, is able to impact on the development of mindreading. More radically, de Villiers and de Villiers (2000, 2003) think that our metarepresentational mentalistic abilities are *constituted* by language; more specifically, the claim is that the mastery of the grammatical rules for embedding tensed complement clauses under verbs of speech or cognition provides children with a necessary representational format for dealing with false beliefs. However, such claim seems to be at odds with the evidence (Carruthers, 2011, pp. 226ff.). Above all, any theorizing on the relation between language and mindreading must come to grips with the results, which have steadily accumulated over the past 14 years, suggesting that infants under age 2 years are already capable of sophisticated mindreading and can attribute to agents not only motivational states (e.g., goals) and epistemic states (e.g., ignorance), but also counterfactual states (e.g., false beliefs) (Baillargeon et al.,

2016; Baillargeon & Scott, 2017). Such evidence knocks out a constitution-thesis *à la* de Villiers.

On the other hand, Meins' view of the relation between language and mentalization is also problematic since the attachment environment is a form of scaffolding that begins with proto-conversational exchange, and only later becomes linguistic. So much so that Meins has recently investigated the caregiver's *nonverbal* appreciation of the infant's mind as reflected in the bodily movements of caregiver and infant ("parental embodied mentalizing"), making the hypothesis that the latter is a behavioral manifestation of mind-mindedness (Shai & Meins, 2018). The results suggest that early parental mentalizing is multifaceted, and multimodal, and that it is useful to explore this interpersonal complexity in verbal and nonverbal, explicit and implicit, and behavioral and representational ways when studying parent-infant relationships.

With these specifications in mind, we can at least claim that when the socio-communicative interaction with caregivers moves from the preverbal to the verbal stage, a whole new range of mature mentalistic activities, which exploit the basic mindreading capacities, emerges under the thrust of caregivers' mind-minded talk. To the extent that these mentalistic activities get turned toward the self, "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" (Gergely & Unoka, 2008, p. 74)[6].

Moreover, from a very young age, a skeletal framework of abstract principles guides infants' sociomoral reasoning. These principles include fairness, harm avoidance, ingroup support (with its corollaries of ingroup care and ingroup loyalty), and authority. Now, psychological and sociomoral reasoning systems work together seamlessly. Thus, for example, if the mindreading system systematically reads other people's behaviors as actions driven by goals, purposes, intentions, the sociomoral-reasoning system determines what actions are obligatory, what actions are permissible, and what actions are impermissible. And then, on the basis of questions such as, "What does that (the mother or the home cat) want to do?", the child begins to ask *also* what her own intentions are, and what her own inner state is. This appropriation

[6] On the distinction between the "proper" and "actual" domains of an evolved cognitive system, see Sperber & Hirschfeld (2004). On the one hand, the specialized system evolved to represent and react to a set of objects, facts and properties; on the other hand, the system actually reacts to a set of objects, facts and properties. The proper domain of mindreading was originally restricted to inferring and representing the causal intentional mental states of other minds only.

of themes that were initially connected only to the reading of others' behaviors is mediated mainly by a learning that is educational, and hence cultural. In other words, it can be supposed that most of the simplest introspections are forms of learning emerging from the verbal stereotypes and rhetoric through which adults rename the intentions of others. A two-year-old child, perhaps because she is frightened by her granny's cat, perhaps as an act of defiance, gives the cat a boot, and here follow the reconstructive judgements about this episode on the part of the adults, which she is invited to internalize: "Bad child! It didn't mean to claw you at all!", or "It scared you, but perhaps the kitty was more scared than you." And so the young child gradually learns -- always internalizing the (hypothetical) names that the adults give to her inner states -- that inside her there are scares, badness, and so on. She understands that these are contingent social expressions, part of social mediations, but also grasps what "information about herself" means.

Thus introspective self-consciousness takes shape in the child in the context of her relationship with the caregiver -- a relationship that is made first of preverbal proto-communicative exchanges, and then of words, descriptions, designations, evaluations of the person. Through such interaction with caregivers (and then with other social partners) children construct their own identity, both *objective* (for others) and *subjective* (for themselves). As Mead (1934) puts it, the basic mechanism for the development of self-consciousness is "the individual's becoming an object to himself by taking the attitudes of other individuals toward himself within an organized setting of social relationships" (p. 225).

For a long time Mead's hypothesis that introspective self-consciousness (identity-for-oneself) takes shape through a process of internalizing the ways in which others see and define us (identity-for-others) was addressed almost exclusively in sociology. It was Sullivan (1953) the first to grasp the significance of the concept of self as developed by Mead, exploiting it in a psychological and more specifically psychodynamic context. It is worth noting how psychology amended a merely sociological construal of Mead's hypothesis. Such a construal has had the drawback of underestimating the complexity, the fatigue, the creative aspects and the risks of the internalization process. By contrast, developmental, social and dynamic psychology have steered Mead's hypothesis onto the right path, making it clear that infants are active creators not only of their structures of relationship with other people, but also of their ways of self-presentation.

3.3. Autobiographical memory, autobiographical reasoning, and narrative identity

With the emergence of autobiographical memory in early childhood and the development of autobiographical reasoning skills in late childhood through adolescence, psychological self-consciousness evolves in the ability to construct a self-narrative as a layer of personality.

The establishment of an autobiographical memory system originates from three complex social-affective-cognitive achievements:

> First, the individual must move from remembering what happened to remembering that this happened to me. [...] Second, the individual must be able to link past experiences to the present; that the self that experienced events in the past is the same self that experiences events in the present (and will be the same self in the future). [...] Third, the individual must be able to create a personal timeline, to construct a coherent chronologically organized sequence of how events followed one another and are linked together in the past and to the present; essentially, the individual must have a chronological biography of self (Fivush, 2011, p. 570).

3.3.1. From remembering what happened to remembering that this happened to me

In the first place, children must be able to represent not only the "what", "where", and "when" of a past event, but also themselves as the subjects who experienced that event. Most of the theories of autobiographical memory development have been cast in terms of explaining infantile amnesia, the phenomenon by which adults cannot recall most of their early childhood experiences. According to Howe and Courage (1993, 1997), before the preschool period, children lack a critical cognitive or social-cognitive framework that would enable them to encode and store memories in such a way that they could later be retrieved as relevant to the self. This framework is self-consciousness as commonly measured in the mirror task of self-recognition. Awareness of self is thus responsible for "kick-starting" autobiographical memory:

> This is because, now that the self has recognizable features (e.g., sensations, feelings), it can serve to organize and structure experiences in memory. Before this, experiences were simply remembered as events that happened, events that

> were only loosely bound in relatively fragmented trace structures. With the advent of self-consciousness, the events that are now being experienced become personalized, in the sense that they are now events that happened to this self, events that happened to "me" (Howe, 2014, p. 552).

Now, we agree with Howe and Courage that the kick-starting factor in the emergence of autobiographical memory is self-consciousness as measured in the mirror self-recognition task. However, we take issue with the authors' construal of the fixed referent as a "cognitive self-concept", because it assumes a too rich interpretation of mirror self-recognition, according to which children's mark-directed behavior is evidential of an *introspective* form of self-consciousness. Accordingly, we agree with those researchers who claim that self-recognition simply rests on bodily self-awareness (Morin, 2010; Brandl, 2016; Saidel, 2016). Our sense of ourselves in time is thus rooted in the onset of a *physical* form of self-describability: the representation of the bodily self constructed in the second year of life acts as a fixed referent around which personally experienced event memories begin to be organized. In James' terms, the Me to which the infant begins to attach episodic memories is the material self.

3.3.2. From a bodily self in time to a psychological one

The second cognitive achievement is the acquisition of the capacity to link past experiences to the present; the child must understand that the self that experienced events in the past is the same self that experiences events in the present (and will be the same self in the future). A series of experiments by Povinelli and colleagues suggests that this understanding emerges only gradually across human childhood.

In the delayed video self-recognition paradigm (a variation of the mirror task), the experimenter is filmed surreptitiously placing a large sticker on the participant's head, during a distractor task. The sticker remains on the participant's head for a period of three minutes, after which time the participant views the original video recording of the sticker placement. Reaching up to remove the sticker from one's head after viewing this recording is taken to indicate the possession of a temporally extended self-representation. The logic here is that only if the participant recognizes that the individual in the recording of this earlier event is the same individual watching the recording in the present

will they recognize that the sticker is on their head here and now, and hence reach up to remove it. In typical development, this task is passed from around 4 years of age.

Povinelli's (2001) interpretation of these findings is that the concept of a temporally extended self emerges at around 4 years as a function of domain-general changes in the child's representational capacities. Following Perner (1991), Povinelli argues that at about 18-24 months of age infants are able to hold in mind a single representation of an event or object (including one's self) while their perceptual system engages with a primary representation (i.e., current reality). This early system of self-representation underlies the capacity to recognize one's self in the mirror: the infants are able to construct and hold in mind a (secondary) representation of the self while they, at the same time, attend to the image reflected in the mirror (a primary representation of the self) and set up a relation between the two. At about 4 years of age, however, children become able to pass the delayed video self-recognition test because they developed the ability to simultaneously entertain various conflicting representations of the same object or event. This ability enables them to hold in mind, at the same time, various conflicting secondary representations of the self and to understand the causal connection between past, present, and future self-states. The new representational ability therefore makes possible the emergence of "an abstract historical-causal self-concept [...] which integrates memories of previously unrelated states of the self into an organized, coherent, and unified autobiographical self-representation" (Fonagy *et al.*, 2002, p. 247).

It can be doubted, however, whether the delayed self-recognition measure is evidence of the emergence of a continuous *psychological* self through time. For if the task is a valid measure of self-awareness as a psychological self, patients with autism spectrum disorder (ASD) should perform badly on it. But they do not: ASD children can recognize themselves in the delayed image as effectively as do 4-5-year-old typically developing children (see Lind, 2010; Dunphy-Lelii & Wellman, 2012). This suggests that recognizing oneself in the delayed video is really evidence of the capacity to establish causal and temporal relations between past and present states of the self, but that the self in question is the *bodily* self and not the psychological one. ASD children, then, appear to possess a coherent representation of their own bodies across time; however, being impaired in mindreading abilities, they cannot make the transition from physical to introspective self-description (Williams, 2010). In the same direction goes a series of studies by Lagattuta and colleagues, which show that it is not until 5 to

6 years of age that the child begins to understand that *psychological* states persist through time and influence current behavior (Lagattuta, 2014).

3.3.3. From linking past to present to creating a narrative identity

For true autobiography, one must move beyond linking past to present to creating a "personal timeline" that temporally organizes the entirety of one's life story. This personal timeline is *narrative identity*.

Over the last three decades Dan McAdams has developed a life-story model of identity at the interface of personality psychology, life-span developmental studies and cultural psychology. Making a synthesis of James' theory of duplex self, Erikson's view of identity, and Henry Murray's research program on the Study of Lives, McAdams (1985) proposed a theory of identity development in which narrative identity is seen as a social-affective-cognitive structure designed to provide that sense of temporal sameness and continuity that Erikson thought to be a defining feature of identity. Around the same time, Katherine Nelson (1989) proposed a theory of early narrative development that has since been associated to McAdams' theory (see for references, McLean & Syed, 2015, p. 2).

Within McAdams' theoretical framework, narrative identity is the internalized and evolving story of the self which integrates the reconstructed past and the imagined future to provide life with some degree of unity, purpose and meaning:

> With respect to the I, the self functions as a unifying process through which subjective experience is synthesized and appropriated as one's own. On the side of the me, the process of appropriating experience as one's own results in a reflexive conception of self (the me that the I constructs), and such a reflexive product may itself express unity and purpose. Identity in the me is the extent to which the me can be arranged (by the I) as a unifying and purpose-giving story. [...T]herefore, the synthesizing I-process creates unity in the me by fashioning a self-defining product that ideally assumes the form of an integrative life narrative (McAdams, 1997, p. 56).

Importantly, McAdams views narrative identity as a *layer of personality*. Within his conceptual framework for conceptualizing the whole person across her life span, narrative identity hinges on two other cognitive layers. The first consists in a small set of broad *dispositional traits* implicated in social life which account for consistencies in behavioral style from one situation to the next and over time.

The second layer consists of a wide range of *characteristic adaptations* (including goals, strivings, personal projects, values, interests, defense mechanisms, coping strategies, relational schemata) which capture more socially contextualised and motivational aspects of psychological individuality. During personality development, people's internalised and evolving life stories are layered over characteristic adaptations, which are, in turn, layered over dispositional traits. And this process of layering may be *integrative*: the process of selfing may succeed in bringing traits, skills, goals, values, and experiences together into a meaningful life story.

3.3.3.1. Narrative self-construction in a naturalistic key

The claim that the type of continuity that connects psychological states across time in an identity-constituting way is specifically narrative in character is typically associated with concerns about *practical* identity, which is personal identity considered in its connection to ethical concerns, as in the case of Locke's theory of personhood. The claim is that we constitute ourselves as "Lockean persons" (i.e., as morally responsible agents) by forming and using autobiographical narratives. The unity of a person is the unity of an autobiographical narrative (cf. Shoemaker, 2019, §2.3).

In some cases, narrative accounts of personal identity are characterised in opposition to the project to amend Locke's relational memory criterion that can be found in psychological continuity theorists such as David Lewis, Derek Parfit, Sydney Shoemaker, and John Perry. Here the question is a *metaphysical* identity question: on what basis should we reidentify a person as numerically the same despite qualitative differences over time or under different descriptions? Answering such a "reidentification" question calls for a criterion of diachronic numerical identity, a criterion of what makes something one and the same thing as itself at different times. However, when the focus shifts from solely metaphysical puzzles about the persistence of complex objects to the relation between identity and practical and evaluative concerns, the question becomes one of "characterization": what characteristics (character traits, motivations, values, mental and bodily capacities and dispositions, emotional attachments, commitments, memories, and so on) make a person the particular person that she is? Such a question concerns identity in the sense of the Eriksonian "identity crisis"; it is a *psycho/social/ethical* identity question (cf. Schechtman, 1996).

According to some proponents of the narrative view, the psycho/social/ethical identity question is importantly related to the metaphysical identity question. In fact, they first build those activities of *self-interpretation* and *self-creation* that are central to our experience of being persons into the kind of continuity that connects person A and person B across time in an identity-constituting way (cf. Korsgaard, 1989). Then, they identify what enables persons to be actively self-interpreting and self-creating agents with the construction of *self-narratives*. In short, "the metaphysical 'glue' that binds person stages together into persons is narrative connectedness, not 'mere' psychological connectedness" (Schroer & Schroer, 2014, p. 460).

The narrative account, however, comes in different forms. Authors such as Macintyre (1984) and Taylor (1989) view the person as a self-interpreting being in a sense inspired by the hermeneutical tradition, namely a tradition that is largely foreign -- or even hostile -- to naturalistic commitments. An empirically-informed narrativist account of personal identity requires a view of self-interpretation as an activity of narrative reappropriation of the products of the unconscious processing -- an activity implemented by apparatuses such as Dennett's (1991) "Joycean machine", or Gazzaniga's (2011) "interpreter module", or Carruthers' (2011) "mindreading system". In this perspective, persons are self-interpreting beings in a sense that is congenial to a view of personal identity in terms of psychological continuity, but fundamentally foreign to the hermeneutical tradition.

3.4. Autobiographical reasoning

Habermas and Bluck (2000) have described the social-cognitive changes that must take place in order for the adolescent to initiate the crafting of the life story that is at the heart of McAdams's theory.

The life story is most completely manifested in entire life narratives as specific, but rare, linguistic products. A more frequent but only partial manifestation of the life story is what Habermas and Bluck (2000) termed "autobiographical reasoning" which is the activity of using "autobiographical arguments" for creating links between personal experiences and other distant parts of one's life, and to the self and its development (Habermas, 2011). This activity is termed "reasoning" to underscore three aspects: the constructive and interpretative nature of the activity, the both cognitive and communicative nature of it, and its normative aspect implied by its appeal to reason and logic.

The term "reasoning" also alludes to the Piagetian cognitive-developmental tradition, which Habermas and Bluck aim to wed to the narrative tradition. In full harmony with Piaget's constructivism, Habermas and Bluck describe the development of the life story in adolescence as the emergence of a new quality, the "global coherence of the life story" (2000, p. 749). To convey the development of the self, up to the present, life narratives not only require the inclusion of various life events and aspects of the self, but also interpretive connections between events and self in order to create a globally coherent story. Global coherence is the narrative feature that differentiates life narratives from mere lists of unrelated memories from one's life.

Autobiographical reasoning involves four socio-cognitive capacities creating four kinds of coherence that are decisive for the overall global coherence of life narratives: (i) the capacity to create temporal orientation by sequential structure and chronology (*temporal* coherence); (ii) the ability to think about the self in abstract terms (i.e., as embodying certain personality traits) and account for changes or developments in the self over time (*causal-motivational* coherence); (iii) the ability to summarize and interpret themes within stories and apply these to one's own life (*thematic* coherence); and (iv) having an awareness of cultural norms regarding the major milestones and events one is expected to experience during the life course.

In an influential review, Habermas and de Silveira (2008) showed that a life narrative begins to emerge in middle childhood, but the coherence of this narrative (in all its dimensions) increases during adolescence. Köber, Schmiedek and Habermas (2015) longitudinally extended this study to explore the development of global coherence in life narratives from childhood to adulthood. It was found that measures of temporal and causal-motivational coherence increase substantially across adolescence up to early adulthood, as does thematic coherence, which continues to develop throughout middle adulthood.

3.4.1. The defensive nature of self-consciousness

Autobiographical reasoning is *constitutive* of narrative identity. It embeds personal memories in a culturally, temporally, causally and thematically coherent life story; thus, the life story format establishes and re-establishes the diachronic continuity of the self. More specifically, autobiographical reasoning

is a mechanism to compensate for *threats of self-discontinuity* (Habermas & Köber, 2015a,b).

Another mechanism that can create self-continuity consists in assimilating memories to the present self-concept (Conway, Singer & Tagini, 2004). The remembered self is systematically distorted by automatically assimilating it to the present self-concept, increasing the similarity between the present and remembered reflected self, in order to maintain conceptual self-sameness. Now, in circumstances of relative stability, assimilating memories to a current self-concept may be sufficient to establish personal sameness in time or personal stability. However, insofar as such mechanism bridges personal change "simply by reducing the perception of change", it cannot "create self-continuity when change is acknowledged" (Habermas & Köber, 2015a, p. 155). In times of biographical upheaval and change, the diachronic continuity of the self can be re-established by autobiographical reasoning through arguments that spell out transformations and their motives.

The construct of autobiographical reasoning thus brings us to the matter of the defensive nature of self-consciousness.

Breaking with a long philosophical tradition that has viewed self-consciousness as a purely cognitive phenomenon, attachment theory and infant research have shown that the construction of affectional life, over the course of infancy and, subsequently, throughout one's entire life, is closely linked to the construction of a subjective identity that is well-defined and accepted as valid. The description of the self that the young child pursues is an "accepting description", i.e., a description that is indissolubly cognitive (as a *definition* of self) and emotional-affectional (as an *acceptance* of self). Children need a clear and consistent capacity to describe themselves in a manner that is fully legitimized by caregivers, socially valid, capable of attracting attention and serving as a base for ceaselessly renewed affectional transactions.

Even adolescent crisis, and together with it the process of social autonomization in post-adolescence, are largely a problem of identity. According to Erikson, the fundamental problem of adolescence lies in discovering how to move from a heteronomous identity to an autonomous self-definition; and this requires an identity synthesis, i.e., a reworking of childhood identifications into a larger, self-determined set of self-identified ideals. In Jamesian terms, the various parts of the material, social and spiritual selves must be organised into "a new pattern that confers upon the Me a unifying and purposeful sense of identity" (McAdams & Cox, 2010, p. 164). The optimal

outcome of such a process is a kind of dialectic balance in which the ego syntonic pole of identity synthesis is predominant over the ego dystonic pole of "identity diffusion". The latter is conceived by Erikson as an insufficient integration of self-images originating from a "weakness of the ego".[7]

This claim leads us into the psychopathological dimension of the inextricable link between identity self-description and self-consciousness. One cannot ascribe concreteness and solidity to one's own self-consciousness if it does not possess at its center, and as its essence, a description of identity that must be clear and, inextricably, "good", in the sense of being worthy of love (Balint, 1965). If the self-description becomes uncertain, the subject soon loses the feeling of being present. The incessant construction and reconstruction of an acceptable and adaptively functioning identity is therefore the process through which our intra- and inter-personal balances are produced, and hence the foundation of psychological well-being and mental health. And this process is the ongoing construction of a system of defenses, the continuously renovated capacity to curb and cope with anxiety and disorder (Jervis, 2014).

This finds illustration in the theories of object relations and attachment, whose theoretical focus is on problems arising from a weakness, fragility, scarce cohesion or insufficient integration of those structures of the mind that Freud calls "das Ich" (essentially, the system of defenses). This structural condition of fragility is experienced by the subject as a chronic feeling of insecurity, or lack of self-esteem, lack of confidence in oneself.

Drawing on Laing (1960), we can describe the experiences originating from a fragility of the ego as symptoms of "ontological insecurity". And in the context of attachment theory Laing's symptoms of ontological insecurity are seen as the last traces of a remote "basic fault" (Balint, 1992), which is to be traced back mainly to early deficiencies in the infant-caregiver relationship. In this context, the idea of an ego that is fragile, or the idea of a self that lacks "cohesion", identifies a condition that predisposes individuals to a broad and varied pathology including psychoses and personality disorders.

4. Conclusion

In this article we outlined a neo-Jamesian approach to the self within a naturalistic, bottom-up, and systemic-relational framework.

[7] McAdams (1997, p. 57) notes that here Erikson takes Freud's Ego as a synthesizing process, and thus coinciding with the selfing process.

In this approach, consciousness of the body as one's own body is a necessary precondition of self-consciousness as psychological self-awareness, and hence of a socially and historically situated narrative self. Thus we take on board the criticism (voiced in different ways by Atkins, 2008; Mackenzie, 2008, 2009; Brandon, 2016) of those accounts of the narrative self that pay little attention to embodiment (e.g., Schechtman, 1996), or go to the extreme of stating that the narrative self is abstract and hence not embodied (e.g., Dennett, 1991). But at the same time, we reject the idea that the embodiment of the narrative self is provided by a *pre-reflective self-consciousness*, a tacit, non-intellectual sense of self that makes every conscious state a first-person phenomenal state and is always both embodied and embedded in the world (Zahavi, 2009; Gallagher & Zahavi, 2019). [8] By contrast, we view self-consciousness as *a construction all the way down*, which develops from automatic and pre-reflective processing of representations of objects (object-consciousness), through *awareness* and then *self-awareness* of the body, up to introspective self-awareness and then narrative identity.

This form of constructivism -- as we have seen, a social-affective-cognitive constructivism -- can be considered as a naturalistic reinterpretation of the Hegelian idea that selfhood is socially constructed and self-experience intersubjectively mediated. However, it enables to move away from that blend of narrativism and epiphenomenism that characterizes some strands of the socio-constructivist approach to the narrative self (for references, see Tekin, 2014). People's self-defining life stories have an intrinsically defensive nature; the description-narration of one's own inner life is organized on the basis of the fundamental need to construct and defend a self-image endowed with an at least minimal solidity. Thus, far from being an epiphenomenal, transient phenomenon, the incessant construction and reconstruction of an acceptable and adaptively functioning identity is the process that produces intra- and inter-personal balances, and hence serves as a foundation of psychological well-being and mental health. The selfing process, therefore, imposes a teleology of self-defense on the human psychobiological system. Accordingly, the storied Me that the selfing process makes is not "empty chatter", but rather a causally efficacious layer of personality viewed as a self-unifying system (cf. Di Francesco, Marraffa & Paternoster 2016, ch. 5).

[8] For a criticism of the interpretation of evidence from research into early social cognitive development in terms of "pre-reflective self-consciousness", see Di Francesco, Marraffa & Paternoster (2016, ch. 3); for a reply, see Gallagher (2017).

REFERENCES

Allen, C., & Trestman, M. (2017). Animal consciousness. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), URL=<https://plato.stanford.edu/archives/win2017/entries/consciousness-animal/>.

Atkins, K. (2008). Narrative identity and embodied continuity. In K. Atkins & C. Mackenzie (Eds.), *Practical Identity and Narrative Agency*. New York: Routledge, 78-98.

Baillargeon, R., Scott, R. M. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*, 237-249.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, *67*, 159-186.

Baillargeon, R., Setoh, P., Sloane, S., Jin, K., & Bian, L. (2014). Infant social cognition: Psychological and sociomoral reasoning. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The Cognitive Neurosciences V*. Cambridge, MA: MIT Press, 7-14.

Balint, M. (1965). Early developmental states of the ego. Primary object love. In Id., *Primary love and psycho-analytic technique*, London: Tavistock, 90-108 (ed. or. 1937).

Balint, M. (1992). *The Basic Fault*. Evanston (Ill.): Northwestern University Press (orig. ed. 1968).

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, *58*, 373-403.

Bateson, M. C. (1979). The epigenesis of conversational interaction: A personal account of research and development. In M. Bullowa (Ed.), *Before Speech: The Beginnings of Human Communication*. Cambridge: Cambridge University Press, 63-77.

Bechtel, W. (2001). The compatibility of complex systems and reduction: A case analysis of memory research. *Minds and Machines*, *11*, 483-502.

Bechtel, W. (2008). *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.

Bechtel, W. & Abrahamsen, A. (2013). Thinking dynamically about biological mechanisms: networks of coupled oscillators. *Foundations of Science*, *18*, 707-723.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge (MA): MIT Press,

Bermúdez, J. (1998). *The Paradox of Self-Consciousness*. Cambridge (MA): MIT Press.

Bowles, S., & Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton and Oxford: Princeton University Press.

Brandl, J. L. (2018). The puzzle of mirror self-recognition. *Phenomenology and Cognitive Science*, *17*, 279-304.

Brandon, P. (2016). Body and self: an entangled narrative. *Phenomenology and the Cognitive Sciences*, *15*, 67-83.

Brownell, C. A., Svetlova, M., & Nichols, S. R. (2012). Emergence and early development of the body image. In C. Brownell & V. Slaughter (Eds.), *Early Development of Body Representation*. Cambridge: Cambridge University Press, 37-58.

Buyukozer Dawkins, M., Ting, F., Stavans, M., & Baillargeon, R. (in press). Early moral cognition: A principle-based approach. In D. Poeppel, G. R. Mangun, & M. S. Gazzaniga (Eds.), *The Cognitive Neurosciences VI*. Cambridge (MA): MIT Press.

Carpendale, J., & Lewis, C. (2006). *How children develop social understanding*. Oxford: Blackwell.

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Carruthers, P. (2019). Questions in development. In L. Butler, S. Ronfard & K. Corriveau (Eds.), *The Questioning Child: Insights from Psychology and Education*. Cambridge: Cambridge University Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge (MA): MIT Press.

Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: replacing scholasticism with science. *Philosophy of Science*, *75*, 1-27.

Chiaradonna, R., & Marraffa, M. (2018). Ontology and the self: ancient and contemporary perspectives. *Discipline filosofiche*, 28, 33-64.

Clark, A. (1997). *Being There*. Cambridge (MA): MIT Press.

Clark, A. (2008). *Supersizing the Mind*. Oxford: Oxford University Press.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*, 594-628.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*, 261-288.

Conway, M. A., Singer, J. A., & Tagini, A. (2004). The self and autobiographical memory: correspondence and coherence. *Social Cognition*, *22*, 491-529.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transaction of the Royal Society B*, *346*, 1149-1157.

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little Brown.

De Villiers, J., de Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell and K. Riggs (eds.), *Children's Reasoning and the Mind*. Hove (UK): Psychology Press, 189-226.

De Villiers, J., de Villiers, P. (2003). Language for thought: coming to understand false beliefs. In D. Gentner & S. Goldin-Meadow (eds.), *Language in Mind*. Cambridge (MA): MIT Press, pp. 335-384.

Di Francesco, M., Marraffa, M., & Paternoster, A. (2016). *The Self and its Defenses. From Psychodynamics to Cognitive Science*. London: Palgrave-Macmillan.

Di Francesco, M., Marraffa, M., & Paternoster, A. (2019a). Self-narrative as a causal center of gravity. *Rivista Italiana di Filosofia del Linguaggio*, 13, 65-75.

Di Francesco, M., Marraffa, M., & Paternoster, A. (2019b). From persons to selfing. Some developments of Parfit's legacy. *Argumenta*, 9, in press.

Dunphy-Lelii, S., & Wellman, H. (2012). Delayed self-recognition in autism: a unique difficulty? *Research in Autism Spectrum Disorders*, *6*, 212-223.

Eagle M.N. (2013). *Attachment and Psychoanalysis: Theory, Research, and Clinical Implications*. New York: Guilford.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion*. Chichester (UK): Wiley and Sons, 45-60.

Ellenberger, H. F. (1970). *The Discovery of the Unconscious*. New York: Basic Books.

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect Regulation, Mentalization, and the Development of the Self*. London: Other Press.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S. (2017). Self-defense: Deflecting deflationary and eliminativist critiques of the sense of ownership. *Frontiers in Psychology*, *8*, 1612, doi: 10.3389/fpsyg.2017.01612.

Gallagher, S., & Zahavi, D. (2019). Phenomenological approaches to self-consciousness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), URL = <https://plato.stanford.edu/archives/sum2019/entries/self-consciousness-phenomenological/>.

Gazzaniga, M.S. (2011). *Who's in Charge? Free Will and the Science of the Brain*. New York: HarperCollins.

Gergely, G. (2002). The development of understanding self and agency. In U. Goswami (ed.), *Blackwell Handbook of Childhood Cognitive Development*. Oxford: Blackwell, 26-46.

Gergely, G. (2007). The social construction of the subjective self: The role of affect-mirroring, markedness, and ostensive communication in self development. In L. Mayes, P. Fonagy & M. Target (eds.), *Developmental Science and Psychoanalysis*. London: Karnac, 45-82.

Gergely, G., Koós, O., & Watson J. S. (2010). Contingent parental reactivity in early socio-emotional development. In T. Fuchs, H. C. Sattel & P. Henningsen (Eds.), *The Embodied Self: Dimensions, Coherence and Disorders*. Stuttgart: Schattauer, 141-169.

Gergely, G., Unoka, Z. (2008). Attachment, affect-regulation and mentalization. In E. L. Jurist, A. Slade & S. Bergner (Eds.), *Mind to Mind: Infant Research, Neuroscience and Psychoanalysis*. New York: Other Press, 50-88.

Gergely, G., Watson, J. S. (1996). The social biofeedback theory of parental affect-mirroring: the development of emotional self-awareness and self-control in infancy. *International Journal of Psychoanalysis*, 77, 1181-212.

Gergely, G., Watson, J. S. (1999). Early social-emotional development: contingency perception and the social biofeedback model. In P. Rochat (Ed.), *Early Social Cognition*. Hillsdale (NJ): Erlbaum, 101-37.

Griffiths, P. E., & Scarantino, A. (2009). Emotions in the wild: The situated perspective on emotion. In P. Robbins & M. Aydede (Eds.), *Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press, 437-453.

Griffiths, P. E., & Tabery, J. G. (2013). Developmental Systems Theory: what does it explain, and how does it explain it? *Advances in Child Development and Behavior*, *45*, 65-94.

Guerini, R., & Marraffa, M. (2017). Lockean persons, self-narratives, and eudaimonia. *Teoria*, 2, pp. 65-80.

Habermas, T. (2011). Autobiographical reasoning: arguing and narrating from a biographical perspective. *New Directions for Child and Adolescent Development*, *131*, 1-17.

Habermas, T., & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin*, *126*, 748-769.

Habermas, T., & de Silveira, C. (2008). The development of global coherence in life narratives across adolescence: temporal, causal and thematic aspects. Developmental Psychology, 44: 707-721.

Habermas, T., & Köber, C. (2015a). Autobiographical reasoning is constitutive for narrative identity: the role of the life story for personal continuity. In K. C. McLean & M. Syed (eds.), *The Oxford Handbook of Identity Development*. Oxford: Oxford University Press, 149-165.

Habermas, T., & Köber, C. (2015b). Autobiographical reasoning in life narratives buffers the effect of biographical disruptions on the sense of self-continuity. *Memory*, *23*, 664-674.

Harré, R. (1986). Mind as social formation. In J. Margolis, M. Krausz, & R. M. Burien (Eds.), *Rationality, relativism and the human sciences*. Dordrecht: Nijhoff, 91-106.

Harré, R. (1987). The social construction of selves. In K. Yardley & T. Honess (Eds.), *Self and Identity: Psychological Perspectives*. New York: Wiley, 41-52.

Hernik, M., Fearon, P., & Fonagy, P. (2009). There must be more to development of mindreading and metacognition than passing false belief tasks. *Behavioral and Brain Sciences*, *32*, 147-148.

Hoerl, C. (2007). Episodic memory, autobiographical memory, narrative: on three key notions in current approaches to memory development. *Philosophical Psychology*, *20*, 621-640.

Holmes, J. (2014). *John Bowlby and Attachment Theory*. London: Routledge (II ed.).

Howe, M. L. (2014). The co-emergence of the self and autobiographical memory. In P. J. Bauer & R. Fivush (Eds.), *The Wiley Handbook on the Development of Children's Memory*. Hoboken (NJ): Wiley-Blackwell, 545-567.

Howe, M. L., & Courage, M. L. (1993). On resolving the enigma of infantile amnesia. *Psychological Bulletin*, *113*, 305-26.

Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review*, *104*, 499-523.

Hutto, D. D. (2008). *Folk Psychological Narratives*. Cambridge (MA): The MIT Press.

Inhelder, B. & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.

Jervis, G. (2001). *Psicologia dinamica*. Bologna: il Mulino.

Jervis, G. (2011). *Il mito dell'interiorità*. Torino: Bollati Boringhieri.

Jervis, G. (2014). *Contro il sentito dire. Psicoanalisi, psichiatria e politica*. Torino: Bollati Boringhieri.

Kant, I. (1998). *Critique of Pure Reason*. Cambridge: Cambridge University Press (orig. ed. 1781-87).

Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations. *Topics in Cognitive Science*, *3*, 438-444.

Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*, 601-627.

Kirk, E., Pine, K., Wheatley, L., Howlett, N., Schulz, J., & Fletcher, B. C. (2015). A longitudinal investigation of the relationship between maternal mind-mindedness and theory of mind. *British Journal of Developmental Psychology*, *33*, 434-445.

Köber, C., Schmiedek, F., & Habermas, T. (2015). Characterizing lifespan development of three aspects of coherence in life narratives: a cohort-sequential study. *Developmental Psychology*, *51*, 260-275.

Korsgaard, C. (1989). Personal identity and the unity of agency: A response to Parfit. *Philosophy and Public Affairs*, *18*, 109-123.

Kouider, S., Stahlhut, C., Gelskov, S., Barbosa, L., Dutat, M., de Gardelle, V., Christophe, A., Dehaene, S., & Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science*, *340*, 376-380.

Kristen-Antonow, S., Sodian, B., Perst, H., & Licata, M. (2015). A longitudinal study of the emerging self from 9 months to the age of 4 years. *Frontiers in Psychology*, *10*(6), 789, doi: 10.3389/fpsyg.2015.00789.

Lagattuta, K. H. (2014). Linking past, present, and future: children's ability to connect mental states and emotions across time. *Child Development Perspectives*, *8*, 90-95.

Laing, R. D. (1960). *The Divided Self*. London: Tavistock.

Leary, M. R., & Tangney, J. P. (2012). *Handbook of Self and Identity* (2nd ed.). New York: Guilford.

Lewis, M. (2014). *The Rise of Consciousness and the Development of Emotional Life*. New York: Guilford.

Lewis, M., & Brooks-Gunn, J. (1979). *Social Cognition and the Acquisition of the Self*. New York and London: Plenum Press.

Lewis, M., & Carmody, D. P. (2008). Self-representation and brain development. *Developmental Psychology*, *44*, 1329-34.

Lichtenberg, J. D. (1989). *Psychoanalysis and Motivation*. Hillsdale (NJ): Analytic Press.

Lind, S. (2010). Memory and the self in autism. A review and theoretical framework. *Autism*, *14*, 430-456.

Lindquist, K., Wager, T., Kober, H., Bliss-Moreau, E., & Barrett, L. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*, 121-143.

Mackenzie, C. (2008). Introduction. In K. Atkins & C. Mackenzie (Eds.), *Practical Identity and Narrative Agency*. New York: Routledge, pp. 1-28.

Mackenzie, C. (2009). Personal identity, narrative integration, and embodiment. In S. Campbell & L. Meynell, & S. Sherwin (Eds.), *Embodiment and Agency*. Pennsylvania: Pennsylvania University Press, pp. 100-125.

Maddie, P. (2001). Naturalism: friends and foes. In J. Tomberlin (Ed.), *Philosophical Perspectives*, *Vol. 15, Metaphysics*. Oxford: Blackwell, 37-67.

Marraffa, M., & Meini, C. (2018). Forms of vitality revisited: The construction of an affective bodily self. *Theory & Psychology*, 29, 27-45.

Marraffa, M., & Meini, C. (2019). A social-affective-cognitive approach to self-construction. From Piaget to Bowlby – and back again. *Paradigmi*, 37, 121-140.

Marraffa, M., & Paternoster, A. (2012). Functions, levels, and mechanisms: The explanation in cognitive science and its problems. *Theory & Psychology*, 23, 22-45.

Marraffa, M., & Paternoster, A. (2016). Disentangling the self. A naturalistic approach to narrative self-construction. *New Ideas in Psychology*, 40, 115-122.

McAdams, D. P. (1985). *Power, Intimacy, and the Life Story: Personological Inquiries into Identity*. Homewood (IL): Dorsey Press.

McAdams, D. P. (1996). Personality, modernity, and the storied self: a contemporary framework for studying persons. *Psychological Inquiry*, 7, pp. 295-321.

McAdams, D. P. (1997). The case for unity in the (post)modern self: a modest proposal. In R. D. Ashmore & L. Jussim (Eds.), *Self and identity. Fundamental issues*. New York: Oxford University Press, pp. 46-78.

McAdams D. P. (2015). *The Art and Science of Personality Development*. New York-London: Guilford Press.

McAdams, D. P., Cox, K. S. (2010). Self and identity across the life span. In R. M. Lerner (Ed.), *The Handbook of Life-Span Development*. New York: Wiley, vol. II, pp. 158-207.

Macintyre, A. (1984). *After Virtue*. Notre Dame, University of Notre Dame Press.

McLean, K. C., & Syed, M. (2015). The field of identity development needs an identity. In Idd. (Eds.), *The Oxford Handbook of Identity Development*. Oxford: Oxford University Press, pp. 1-10.

Mead, G. H. (1934). *Mind, Self, and Society*. Chicago: University of Chicago Press.

Meins, E. (2011). Social relationships and children's understanding of mind: attachment, internal states, and mind-mindedness. In M. Siegal & L. Surian (Eds.), *Access to language and cognitive development*. Oxford-New York: Oxford University Press, 23-43.

Meins, E. (2013). Sensitive attunement to infants' internal states: operationalizing the construct of mind-mindedness. *Attachment & Human Development*, *15*, 524-544.

Meins, E., Fernyhough, C., de Rosnay, M., Arnott, B., Leekam, S. R., & Turner, M. (2012). Mind-Mindedness as a multidimensional construct: Appropriate and nonattuned mind-related comments independently predict infant-mother attachment in a socially diverse sample. *Infancy*, *17*, 393-415.

Meltzoff, A. N. (2013). Origins of social cognition. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World*. Oxford: Oxford University Press, 139-144.

Mitchell, S. A. (1988). *Relational Concepts in Psychoanalysis*. Cambridge (MA): Harvard University Press.

Morin, A. (2010). Self-recognition, Theory-of-Mind, and self-awareness: what side are you on? *Laterality*, *16*, 367-383.

Musholt, K. (2015). *Thinking about Oneself. From Nonconceptual Content to the Concept of a Self*. Cambridge (MA): MIT Press.

Nelson, K. (1989). *Narratives from the Crib*. Cambridge (MA): Harvard University Press.

Nielsen, M., Dissanayake, C., & Kashima, Y. (2003). A longitudinal investigation of self-other discrimination and the emergence of mirror self-recognition. *Infant Behavior & Development*, *26*, 213-226.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge (MA): MIT Press.

Povinelli, D. J. (2001). The self: elevated in consciousness and extended in time. In C. Moore & K. Lemmon (Eds.), *The Self in Time: Developmental Perspectives*. Mahwah (NJ): Erlbaum, 75-95.

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Quine, W.V.O. (1981). *Theories and Things*. Cambridge (MA), Harvard University Press.

Rochat, P. (2012). Primordial sense of an embodied self-unity. In V. Slaughter & C. Brownell (Eds.), *Early Development of Body Representations*. Cambridge: Cambridge University Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145-172.

Ryan, R. M. (1995). Psychological needs and the facilitation of integrative processes. *Journal of Personality*, *63*, 397-427.

Saidel, E. (2016). Through the looking glass, and what we (don't) find there. *Biology and Philosophy*, *31*, 335-352.

Schechtman, M. (1996). *The Constitution of Selves*. Ithaca (NY): Cornell University Press.

Shai, D., & Meins, E. (2018). Parental embodied mentalizing and its relation to mind-mindedness, sensitivity, and attachment security. *Infancy*, *23*, 857-872.

Schroer, J.W., & Schroer, R. (2014). Getting the story right: a reductionist narrative account of personal identity. *Philosophical Studies*, *171*, 445-469.

Shoemaker, D. (2019). Personal Identity and Ethics, in E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/win2019/entries/identity-ethics/>.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, *8*, pp. 40-46.

Sullivan, H. S. (1953). *The Interpersonal Theory of Psychiatry*. New York: Norton.

Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge (MA): Harvard University Press.

Tekin, Ş. (2014). Self, philosophical considerations. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology*. https://doi.org/10.1002/9781118625392.wbecp519.

Van Essen, D. C., Donahue, C., Dierker, D.L., & Glasser, M.F. (2016). Parcellations and connectivity patterns in human and macaque cerebral cortex. In Kennedy, H., Van Essen, D.C., Christen, Y. (Eds.), *Micro-, Meso- and Macro-Connectomics of the Brain*. Berlin: Springer, 89-106.

Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, *92*, 345-81.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, pp. 1-14.

Zahavi, D. (2009). Is the self a social construct? *Inquiry*, *52*, 551-573.

Williams, D. (2010). Theory of own mind in autism. Evidence of a specific deficit in self-awareness? *Autism*, *14*, 474-494.

Wilson, R. A. (2004). *Boundaries of the Mind. The Individual in the Fragile Sciences: Cognition*. Cambridge: Cambridge University Press.