



# 'I Interact Therefore I Am': The Self as a Historical Product of Dialectical Attunement

Dimitris Bolis<sup>1,2</sup> · Leonhard Schilbach<sup>1,2,3,4</sup>

© The Author(s) 2018

## Abstract

In this article, moving from being to becoming, we construe the 'self' as a dynamic process rather than as a static entity. To this end we draw on dialectics and Bayesian accounts of cognition. The former allows us to holistically consider the 'self' as the interplay between internalization and externalization and the latter to operationalize our suggestion formally. Internalization is considered here as the co-construction of bodily hierarchical models of the (social) world and the organism, while externalization is taken as the collective transformation of the world. We do not consider these processes as sequentially linked, but rather as a dialectic between the collective and the individual. This leads us to the suggestion of the self as a historical product of dialectical attunement across multiple time scales, from species evolution and culture to individual development and everyday learning. Subsequently, we describe concrete means for empirically testing our proposal in the form of two-person psychophysiology and multi-level analyses of intersubjectivity. Taken together, we suggest that a fine-grained analysis of social interaction might allow us to reconsider the 'self' beyond the static individual, i.e. how it emerges and manifests itself in social relations. Such an approach, we believe, could be relevant in multiple fields, from ethics and psychiatry to pedagogy and artificial intelligence.

**Keywords** Social interaction · Intersubjectivity · Internalization · Externalization · Predictive processing · Active inference

*Through others, we become ourselves.*  
*Lev Vygotsky (1896–1936)*  
*I see nothing other than becoming.*  
*Heraclitus (ca. 535–475 BC) in*  
*Nietzsche's the birth of tragedy*

## 1 What is the 'Self'?

### 1.1 Questioning the Question

In modern societies people tend to consider the idea of the 'self' as self-evident. Certain civilizations have even suggested that one's ultimate goal in life is to 'know thyself' (e.g., the ancient Greeks referred to it in multiple instances as 'γνώθι σεαυτόν'). However, the question of the 'self' did not exist from the beginning of the history of culture and human thought, but it arose at a certain level of historical development as a result of deep societal transformations. At different stages of historical development, this question has been addressed in different ways. For instance, Plato (429–347 BC), and before him Homer (ca. Eighth century BC), imagined the 'self' as an immaterial spiritual substance (i.e., the psyche or the soul). More specifically, Plato contrasted the eternal *form* with the ephemeral *body*, which he thought of as an imperfect copy of the former (Kraut 2017). In fact, we later meet dualistic views on the 'self' in various religious traditions, as well as in notable thinkers, such as Platinus (ca. 204–270) and Descartes (1596–1650).

---

✉ Dimitris Bolis  
dimitris\_bolis@psych.mpg.de

Leonhard Schilbach  
leonhard\_schilbach@psych.mpg.de

<sup>1</sup> Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich-Schwabing, Germany

<sup>2</sup> International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany

<sup>3</sup> Department of Psychiatry, Ludwig Maximilian Universität, Munich, Germany

<sup>4</sup> Outpatient and Day Clinic for Disorders of Social Interaction, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich-Schwabing, Germany

Descartes, who famously declared “I think, therefore I am” (or “I doubt, therefore I think, therefore I am”, as paraphrased by Antoine Léonard Thomas), considered *mind* and *body* as two distinct entities, which could yet influence each other.

Nowadays, mainstream science has moved away from an idealistic and dualistic view of the ‘self’. Already, Aristotle had argued that the soul could not be separated from the body (cf. Sihvola 2008). Yet, religion, offering the concept of immortality as a solution to the problem of death has played a pivotal role in hindering this transition (Barresi and Martin 2011), ignoring alternative solutions such as the one put forward by Epicurus (341–270 BC), who proclaimed that the problem is not death itself, but the fear of death. In fact, Epicurus and others such as Democritus adopted a monistic perspective, which can be traced into modern times with thinkers such as Pierre Gassendi (1592–1655), Baruch Spinoza (1632–1677) and Ludwig Feuerbach (1804–1872). Despite other fundamental differences, present in diverse philosophical traditions has been the idea of a lawful understanding of the world, at times emphasizing a mechanistic explanation, which largely characterizes the scientific paradigm until today. Indeed, one can draw parallels in today’s neuroscience, which is largely grounded in frameworks, which focus on describing the underlying mechanisms of a phenomenon, e.g., distinct neurobiological mechanisms underlying consciousness.

Various roles have been considered for the ‘self’ and consciousness in (more) modern science and philosophy as well. For instance, John Locke (1632–1704) focused on the relations between basic physical/mental elements, emphasizing sameness: “[...] in this alone consists personal identity, i.e. the sameness of a rational being: And as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done.” (Locke 1694; cf. Barresi and Martin 2011). On the other hand, David Hume (1711–1776) claimed that the ‘self’ is an illusion, as there “are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is compos’d” (Hume 1739; cf. Barresi and Martin 2011). After all, Friedrich Nietzsche (1844–1900) not only famously argued that “God is dead”, but also noted that the ‘self’ is dead as well (cf. Barresi and Martin 2011). So, what is the ‘self’? Barresi and Martin (2011) answer that the concept of ‘self’ in today’s literature appears divided in a number of different roles, such as ‘self-image’, ‘self-conception’, ‘self-discovery’, ‘self-confidence’ etc.

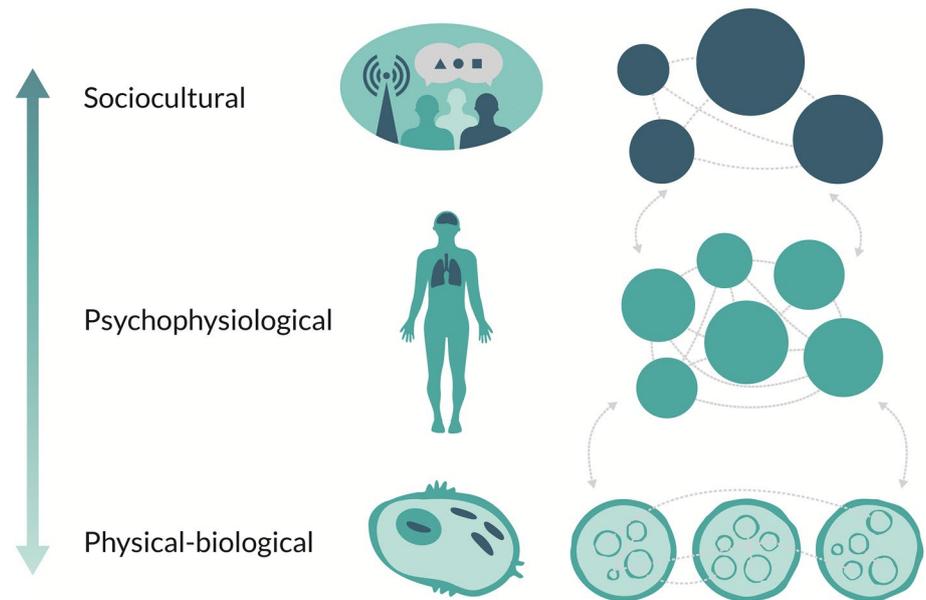
In this article, we will approach the multi-fragmented paradox of the self, through an integrative perspective,

adopting a dialectical and historical perspective. In line with dialectical cultural-historical theories (cf. Vygotsky 1930–1935/1978), we will try to motivate a shift of focus from *being* to *becoming*, along multiple temporal scales. In doing so, we will move beyond the individual in the question of the (a-)typical ‘self’, in both conceptual and empirical regards. More concretely, we will argue that the ‘self’ lies beyond the static individual, namely in the *unfolding of social relations, as a dialectic of internalization/externalization, over multiple temporal scales*. Along these lines, autism and other psychiatric conditions have been recently revisited as processes of cumulative *misattunement between persons*, rather than mere brain disorders (Bolis et al. 2016, 2017). Subsequently, we will delineate an empirical research framework for scientifically validating relevant questions, i.e. two-person psychophysiology and multi-level analyses of intersubjectivity. Finally, putting this approach into a broader context, we will discuss why challenging the concept of the self is important anyway by describing the practical implications of our approach across various fields of research and practice. Here, we will consider aspects ranging from ethics and pedagogy to psychiatry, neuroscience and artificial intelligence.

## 1.2 A Dialectical Perspective

To begin, we will make a case for the use of dialectics as a powerful tool for science. To this end, we will first provide a brief introduction to the method and present concrete dialectical insights for the discussion of the self. Dialectics can be thought of as an evolving school of thought, met in various historical and cultural contexts (e.g., Greek, Chinese, Indian, German dialectic; Wong 2006; Dafermos 2015). It asserts that phenomena cannot be meaningfully understood by reducing them into single levels of description or by assuming a metaphysical independence between levels of description. It rather states that phenomena should be studied as processes in their wholeness, inner contradiction and movement. In this light, the self cannot be understood in isolation from the body, social interaction and society (Bolis et al. 2017). More concretely, primarily leaning on views of Vygotsky and colleagues on the dialectical nature of human thought and development, we will try to overcome traditional dichotomies, such as object/subject and organism/environment, by viewing them as both a result and a cause of reciprocal adjustments, or individual/society by considering the whole and the part as, albeit partially autonomous, highly interdependent levels of organization. Along these lines, the self is not to be taken as a static entity bounded by the individual, but rather as the interplay of dynamically and reciprocally interacting factors. More specifically, we will consider it as a process of circular causality among different levels of organization (Fig. 1; e.g., *physical, biological,*

**Fig. 1** Schematic depiction of dynamic interrelationships in the evolution of matter organization across several time scales



*psychophysiological and sociocultural*) unfolding over different time frames (e.g., *evolutionary, cultural, developmental, psychophysiological and microbiological scales*; Vygotsky 1930–1935/1978; Bolis et al. 2017).

In a nutshell, dialectical thought emphasizes change over sameness and becoming over being, by viewing reality as dynamic processes rather than static entities (see also process philosophy; Seibt 2017). As Nietzsche (1844–1900) noted, citing Heraclitus (ca. 535–475 BC): *he [Heraclitus] altogether denied being. [...] Louder than Anaximander, Heraclitus proclaimed: “I see nothing other than becoming. Be not deceived. It is the fault of your short-sightedness, not of the essence of things, if you believe you see land somewhere in the ocean of becoming and passing-away. You use names for things as though they rigidly, persistently endured; yet even the stream into which you step a second time is not the one you stepped into before”* (Nietzsche 1872/1999, pp. 51–52). Along these lines, Nietzsche strictly denies a dichotomy of object/subject: “[the subject is but a] term for our belief in a unity underlying all the different impulses of the highest feeling of reality”. There is no such unity, only “the fiction that many similar states in us are the effect of one substratum: but it is we who first created the “similarity” of these states; our adjusting them and making them similar is the fact, not their similarity, which ought rather to be denied.” (Nietzsche 1901/2017; see also Barresi and Martin 2011). Nietzsche then goes on to criticize an “absurd overestimation of consciousness” which had been transformed “into a unity, an entity: ‘spirit’, ‘soul’, something that feels, thinks, wills”, provocatively characterizing this as one of the “tremendous blunders” intellectual culture had created (Nietzsche 1901/2017; see also Barresi and Martin 2011). In other words, Nietzsche here rejects the idea of an ‘artificial’

unity of consciousness (or the self). This brings us to a cardinal concept of dialectics, the ‘unity of opposites’.

Put simply, ‘unity of opposites’ defines a phenomenon by its internal oppositions: “All things come into being by conflict of opposites, and the whole flows like a stream” (Diogenes on Heraclitus, ca. Third century BC; cf. Magnus 1970). Later Hegel (1770–1831) elegantly elaborated: “[...] every actual thing involves a coexistence of opposed elements. Consequently to know, or, in other words, to comprehend an object is equivalent to being conscious of it as a concrete unity of opposed determinations, [whereas] the old metaphysic, as we have already seen, when it studied the objects of which it sought a metaphysical knowledge, went to work by applying categories abstractly and to the exclusion of their opposites” (cited in Blunden 2000). In brief, Hegel claimed that ideas and concepts can be only understood in historical terms, as when abstracted, they become meaningless (Grossmann 2018). Importantly, a dialectical account does not merely focus on interpreting a harmonic development of internal contradictions, but also unveils dramatic tensions, conflicts and struggle of opposites. In fact, within dialectical thinking, such inner contradictions are the ones that drive change. Gradual change, in turn, is thought to lead to ‘crises’, which are overcome by qualitative leaps. Taken together dialectics, therefore, assume a constant evolution of phenomena, where change is periodic but not returning to the same point.

With regard to the topic at hand, Hegel suggested that self-consciousness does not emerge through passive and individualistic introspection, but through dynamic and reciprocal relations with others (cf. Barresi and Martin 2011). In fact, Karl Marx (1818–1883), whose work leaned on but also criticized Hegel’s work, proclaimed: “[...] the human

essence is no abstraction inherent in each single individual. In its reality it is the ensemble of the social relations” (Marx and Engels 1888). The primacy of the social realm has later been stressed by so-called cultural-historical approaches (cf. Roth 2016). A prime example of this can be found in the work of Lev Vygotsky (1896–1934), who directly applied dialectical thinking to developmental psychology and proclaimed that “through others we become ourselves”. He further suggested that all ‘higher’ mental processes *within* an individual result from an internalization of prior social interactions *between* people. But dialectical thinking, as described here, should not be exclusively attributed to Western philosophy. For instance, according to African Ubuntu “a person is a person through other persons” (Birhane 2017). We also meet forms of dialectical thinking in Buddhism and Taoism (cf., Grossmann 2018). Taken together, in the formation of the self, *the social can be assumed to dialectically precede the individual*.

In this line of thought, we suggest that *interpersonal statistical regularities shape multiscale hierarchical models on an individual level and vice versa*. For instance, at the level of perceptual awareness and everyday learning (time scale of seconds to hours), others play an important role in shaping subjective feelings and decision-making. Let us imagine an illustrative scenario (Bolis and Schilbach 2017b): a person, in the process of deciding what is the most appropriate clothing for tonight’s walk, checks current weather out at the balcony. She feels a cold breeze, which initially makes her think that a warm coat might be a good idea; yet a glance down the road makes her change opinion, as all people outside this day are lightly dressed. Such kind of decisions, especially when reinforced by persistent cultural norms (time scale of weeks to years), is possible to form even more stable personal habits. For instance, people in ancient Greece were accustomed to exercising without clothes. In modern societies, despite objective conditions that might call for such a habit sometimes (e.g., warm weather) such a behavior would be considered uncomfortable by most people. Here we see how a socially constructed statistical regularity is internalized at the level of the individual—at such an extent so that its violation directly evokes certain subjective feelings.

Across longer time scales, the cumulative internalization of such interpersonal regularities directly shapes who we are becoming, literally changing our bodies and brains. Both ‘higher level’ mental functions and ‘automatic’ processes can be thought of as emerging due to and through social interactions across the life span. Let us examine a simple example inspired by Vygotsky (1930–1935/1978, pp. 103 of the Greek translation): a child in an effort to maintain interoceptive balance unsuccessfully tries to reach for food with the index finger extended. The caregiver, who observes the effort, helps with bringing the food toward the hand of the child. After a number of repetitions, this kind of

interpersonal process, and the statistical regularities associated with it, is internalized by the child in such a way that the extension of the index finger eventually represents a call for attention to a pointed object. Intriguingly, it is not only higher symbolic functions that are culturally shaped, but also more ‘fundamental’ ones, such as eating and drinking. For instance, while babies eat and drink when they feel the need for it, adults regularly do so not for covering direct survival needs but rather social ones (e.g., eating as a part of a break from work or drinking alcohol when socializing). Along these lines even interoceptive control can be thought of having social origins, being developed in this way already from infancy onwards (Ciaunica and Fotopoulou 2017; Fotopoulou and Tsakiris 2017). To probe this further, interaction *processes* can be thought of as ontologically primary to *entities* on an ultimately basic level. In this line of thinking, entities actually emerge through interactions (intra-actions for emphasis; Barad 2003) and not vice versa, from within their relationship and not outside. In this light, we view the multifaceted construction of the self as an active process of culturally mediated internalization of social interactions along multiple time scales.

Here, it is crucial to note that internalization plays an important albeit partial role in the formation of the self. It is the dialectic between internalization and externalization that provides a more complete picture of the co-construction of individual and social reality. Internalization can be thought of as the active reconstruction and synthesis of incoming information and past experiences, while externalization can be thought of as the tool-mediated translation of inner processes into collectively transforming the world, including others. The dialectic between internalization and externalization becomes apparent when examining the simple example of holding and manipulating an object (Leontyev 1975/1983; Stetsenko and Arievitch 2004). In this very moment a person transforms not only the world but also *her-self*, as in her effort to act on the environment, she embodies its structure and dynamics. Tools are not to be confused only with conventional material objects. The term here is used to also broadly encompass ‘intellectual objects’ in the service of communication, such as language and art (cf. Vygotsky 1930–1935/1978; Dafermos 2002). In other words, humans change themselves through changing the environment in a socioculturally mediated procedure of mutual adjustment. More broadly, evolution (or development<sup>1</sup>) of species, societies, persons and concepts should not be viewed as an one-way adjustment, but rather as a dialectical, namely a dynamic, reciprocal and cumulative process (cf. Levins and Lewontin 1985). We will come back to this crucial insight

<sup>1</sup> Hereafter we will use the term development to broadly imply change across various scales, emphasizing our historical standpoint.

later, but will now review these cardinal concepts through a Bayesian lens, which will allow us to operationalize our suggestions formally.

### 1.3 A Bayesian Perspective

The main premises of the “Bayesian brain hypothesis” rest on the idea that the brain represents information accessed via the sensory organs in the form of probability densities, as opposed to single numbers, which are continuously updated, as if following a specific set of mathematical formulas based on Bayes theorem (cf. Bolis et al. 2017). Interestingly, such a perspective brings together under a common umbrella diverse putative cognitive processes of major importance, such as optimal information integration both in time and space, optimal multimodal cue integration, as well as flexible information manipulation without the need to commit to particular decisions at an early stage of processing (Knill and Pouget 2004). In other words, through a Bayesian lens one can view the brain as an organ which calculates and maintains probabilities about events in the world or about the organism itself, via a combination of already gained experience and newly sensed information. Importantly, the more confidence (i.e., precision) is placed on the validity of experience (i.e., prior beliefs) the less beliefs are updated based on new incoming information (i.e., evidence). Notably, a Bayesian ‘belief’ should not be confused with an everyday meaning of the word belief which might be taken to refer to a conscious representation. On the contrary, a Bayesian belief can be thought of as a dynamic state, either conscious (e.g., determination not to eat meat) or unconscious (e.g., glucose levels).

A concrete and prominent implementation of the Bayesian brain hypothesis can be found in predictive processing (i.e., predictive coding and active inference; Friston 2010, 2013; Clark 2013). Within this framework a biological system is essentially viewed as a prediction machine and action generator, which actively tries to align reality with internalized models of reality, as precisely as possible. As noted above, reality embraces both the world and the organism itself. According to such a perspective, the brain’s ultimate goal is the long-term minimization of free energy, by calculating (under certain simplifying assumptions) *prediction errors*, i.e., the discrepancy between incoming information and generated predictions, based on prior experience. Importantly, this is thought to be accomplished via two main avenues, namely either via updating the (Bayesian) beliefs one holds for aligning them with the environment (i.e., predictive coding; cf. internalization), or through action, which can help to experience the environment in accordance with prior beliefs (i.e., active inference; cf. externalization). Put simply, to survive, an organism obeys the following straightforward

rule: *adjust yourself to reality or change the reality itself* (cf. Friston et al. 2010).

In this framework, the updating of beliefs is accomplished across various hierarchical levels at the same time. More concretely, two processes run in parallel: prediction errors ascend the hierarchy reconfiguring the organism for optimizing predictions, while in parallel predictions descend the hierarchy explaining away prediction errors. The hierarchical organization of this scheme is of immense importance, as it allows for the consideration of multiple levels of increasing abstraction. For instance, social relations along development are not merely stored and represented as concrete memories, but are perhaps more crucially, internalized at higher levels of the hierarchy as generalized cultural norms. The latter can, in turn, be utilized to guide behavior across a multitude of contexts.

As noted above, a process of belief updating should be always thought of in relation to action. Importantly, such a dialectic of internalization and externalization can take either ‘adaptive’ or ‘maladaptive’ forms along various time scales, leading to a cascade of interpersonal (mis-)attunement (Bolis et al. 2017). To give a simple example, abusive interactions with care-givers in early life could influence the way an individual forms relations later, which may help to explain personal tendencies and so-called personalities, but also symptoms across different psychiatric and psychological conditions. In other words, growing up in an interpersonally adverse environment may lead to expectations about how social interactions unfold, which will modulate how future interactions actually play out. From our standpoint, such an example illustrates how the Bayesian perspective may be able to capture and express the inextricable linkage of social and individual reality. Seen through a Bayesian and dialectical lens at the same time, we can, therefore, view the ‘self’ as a non-linear dynamic *process*, rather than as a static and unified *entity*.

Notably, predictive coding and active inference can be thought of as a dialectical framework in and of itself. Perception and action become two dialectical facets of the same process, i.e. the minimization of prediction error. Current internal (e.g., perceptual) states inform future actions, while informed interaction with the environment (including others) greatly modulates internal states. Furthermore, the interrelation between the environment and the ‘self’ is controlled by the synthesis of an organism’s current state and incoming information, either through updating current beliefs or the environment itself. In these terms the ‘self’ can be considered as the dialectic between predictive coding (cf. internalization) and active inference (cf. externalization) processes (Bolis et al. 2017). Taken together, multilevel computational frameworks grounded in predictive processing (cf. Bolis and Schilbach 2017b; Ramstead et al. 2017) can, therefore, serve as a formal bridge between philosophical arguments

and neuropsychological evidence for revisiting the ‘self’ as a historical product of dialectical attunement.

## 2 The Dynamic Self in Action

### 2.1 The Dialectic of Internalization/Externalization: Insights from Evolutionary & Developmental Psychology, Neuroscience & Psychiatry

In the following section, we selectively review results and insights from different disciplines in order to add empirical findings to the argument that the self can be regarded as a (historical or developmental) dialectic of internalization/externalization over multiple scales.

Across an evolutionary scale, the change to upright position comprises perhaps one of the most important qualitative leaps. In fact, bipedal walking has been crucial to the evolution of the self for various reasons. Perhaps, most importantly, walking on two feet allowed the development and use of sophisticated tools. The latter revolutionized the way humans adapt to the environment, allowing them to actively and dialectically transform the world they inhabit according to their needs. That is, it is not only humans who change the environment, but the environment in turn changes them in face of their impact on it (cf. Levins and Lewontin 1985). In brief, contrary to a perhaps common belief, humans (and other organisms) do not evolve via passive adaptation, but they fundamentally change themselves via socioculturally mediated transformations of the environment. However, having said that, this development has not come without compromises.

It has been hypothesised that bipedal walking has imposed certain constraints on the birth canal, which does not allow the birth of a fetus much older (and thus bigger) than 9 months. Additionally, according to the ‘metabolic crossover hypothesis’ (Ellison 2001; Dunsworth et al. 2012) the mother may not be able to support an older and more energetically demanding fetus. Consequently, while apes and other animals quickly master basic skills that grant them relatively early independency after birth, human infants are born unable to survive on their own. Indeed, the brain size of newly born infants is only a quarter of its fully developed size. This means that major development occurs after birth in direct interaction with the environment and others: “Maybe human newborns are adapted to soaking up all this cultural stuff and maybe being born earlier lets you do this [...] Maybe being born earlier is better if you’re a cultural animal” (Karen Rosenberg on Adolf Portman; cited in Wong 2012). Such a compromise between early independency and optimal development might actually, in and of itself, define the timing of birth.

Another major evolutionary leap with regards to human cognition is the change from individual to shared intentionality (Vygotsky 1930–1935/1978; Tomasello 1999, 2014; Tomasello et al. 2005; Tomasello and Carpenter 2007), which can be broken down to more intermediate leaps (e.g. from individual to joint and from joint to collective intentionality; Tomasello 2014). The question here is: How did we go from relatively competitive great ape societies to (possibly) cooperative human cultures? It might have been a huge leap if there had not been an intermediate link between our common ancestor and humans. The needs for cooperation (e.g. for foraging) in the early human societies may have led to the transformation of individual to joint intentionality, involving two (or a small number of) individuals (Tomasello 2014). According to this hypothesis, this development has allowed for the coordination of roles and perspectives toward joint objectives, resulting in new forms of perspectival and symbolic representations, socially recursive inference and self-monitoring (regulating one’s own actions from the perspective of a cooperative partner). The practical need for coordination might have actually prompted the development of bodily structures, which subsequently supported more abstracted cognitive functions beyond the ‘here and now’. One tempting line of thought here would be to consider human body (e.g., eye and face) and brain evolution as reciprocally driven in the context of collaborative social interaction (cf. Kobayashi and Kohshima 2001; Powell et al. 2010; Dobson 2012). From a Bayesian perspective, ascending the hierarchy of a neural network, information gets more and more abstracted (e.g., from dealing with the probability of an event, to dealing with volatility, volatility of volatility and so forth; cf. Mathys et al. 2011). Taken together, we hypothesize that such a kind of evolution, which have allowed for abstracting beyond the concreteness of real-time social interactions, might have been toward the direction of extended bodily hierarchies.

Similarly to development at the scale of phylogeny, development at the scale of ontogeny can be also thought of as unfolding in socioculturally mediated interaction with the environment and others, undergoing a series of qualitative leaps along the lifespan (e.g., from individual to collective intentionality; Tomasello and Rakoczy 2003). More concretely, the acquisition of language, which can be considered as a particularly transformative leap for social cognition and interaction, is thought to emerge out of various pre-speech communicative acts (cf. Bruner 1974). An initial basic form of dyadic interaction (between the infant and the caregiver) could serve as the substratum for the development of joint attention, as well as more complex forms of interaction. For instance, dyadic (face-to-face) and triadic (including an object) interactions have found to be developmentally linked (Striano and Rochat 1999). Furthermore, joint attention, which is observed before fully developed social-cognitive

awareness (Brooks and Meltzoff 2005), can predict future linguistic ability (Morales et al. 2000; Mundy et al. 2007). Additionally, maternal sensitivity (Hobson et al. 2004) and synchronicity (Carpenter et al. 1998) have found to correlate with infants’ propensity to engage in social interactions and language development respectively.

Also in so-called psychiatric disorders, here thought of as disorders of social interaction or cases of so-called atypical social interaction, we find an interrelation between the manifestation of the organic condition and interpersonal difficulties (Vygotsky 1930–1935/1978; Schilbach 2016; Bolis et al. 2017). When it comes to autism, synchronicity in earlier play interactions between the child and the caregiver was found to correlate with the development of subsequent communicative forms, such as joint attention and language (Siller and Sigman 2002). In fact, it has been suggested that autism can be viewed not as a mere brain disorder, but rather as an evolving interpersonal misattunement encompassing various levels of description (Bolis et al. 2017). An attunement between the child and the caregiver along development is crucial in language acquisition. Yet, even when an autistic individual becomes able to talk, in most of the cases they achieve a propositional attunement (knowing that), as opposed to a pragmatic attunement (knowing how), a fact which largely prevents an intuitive participation in interactions with others. This alone, we suggest, might have direct implications in the formation of the self in autism due to the crucial dialectical nature of language.

Our discussion on tool mediated evolution holds also for individual development: language can be viewed as a communicative tool used for transforming the (social) world, but also the self itself (Vygotsky 1934/1962). This dialectical nature of language becomes evident when examining its dual role, in speech (interpersonal) and thought (intrapersonal), which should be thought in unity, rather than in external (even tight) relation (Vygotsky 1934/1962). In other words, contrary to a common assumption that speech is merely an enacted thought, speech and thought unfold together, inextricably entangled. Indeed, recent evidence demonstrates neural coupling during production and comprehension of real-life speech (Silbert et al. 2014). Importantly, the interpersonal aspect of language should be still thought of as temporally and conceptually preceding the intrapersonal one. That is, in contrast to a Piagetian perspective, we adhere to the Vygotskian idea that it is social interaction that drives development and not *vice versa*.

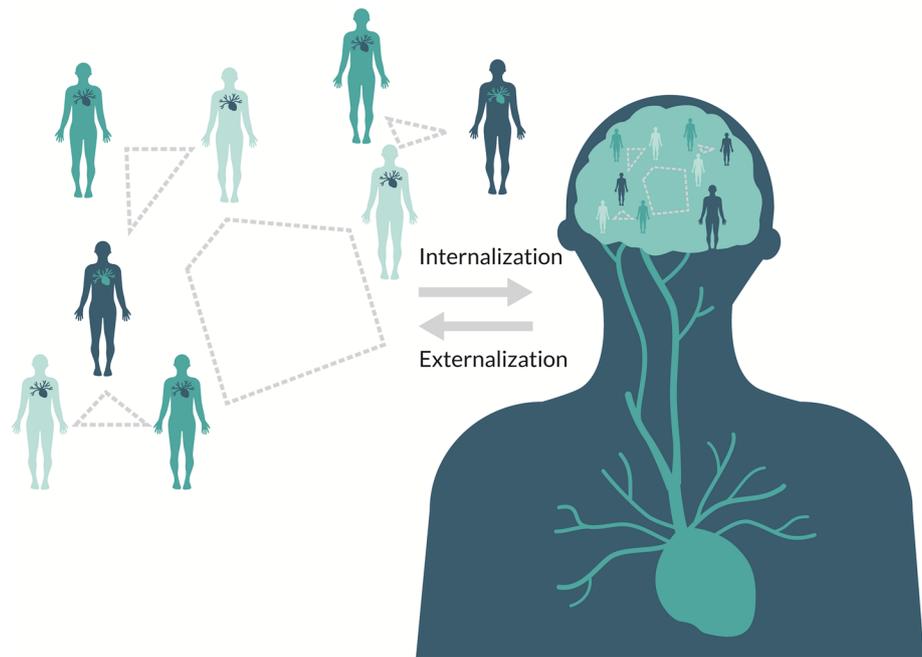
In sum, basic forms of interpersonal sensorimotor contingencies gradually evolve into more complex forms of interactions, such as joint attention and multi-person interactions. This kind of initial social interactions might be exactly what (reciprocally) drives development of social cognition for dealing with beyond ‘the now and here’ (cf. Theory of Mind; Baron-Cohen 1991; Tomasello 1995). At the neural

level, it has been suggested that joint attention might be the outcome of two interacting systems, namely the *posterior* and the *anterior* attention system (Mundy and Newell 2007). The posterior system, which is relatively involuntary and common to many primates, begins to develop during first months of life and can be, simply speaking, thought of as serving for an understanding of “where others’ eyes go, their behaviour follows” (Jellema et al. 2000; Mundy and Newell 2007). The anterior system, which is considered volitional and goal-directed, develops later and can be, along similar lines, thought of as serving an understanding of “where my eye’s go, my behaviours follows” (Mundy and Newell 2007). We take this as suggestive of a claim that the ‘self’ develops tightly connected to the understanding of the ‘other’ and that in fact the latter proceeds.

It might actually be the case that it is exactly in our effort to understand others that we develop an understanding of ourselves. Here, three tangled modeling loops are considered: (i) the inner loop, dealing with the prediction of internal bodily processes (cf. interoception), (ii) the perception–action loop, which involves the anticipation of the consequences of one’s actions on the world and (iii) the self–other loop, which deals with modelling other minds (Timmermans et al. 2012). Exactly the latter loop, through social interactions, might be what ontogenetically forge sophisticated bodily structures that are later deployed for reflective social cognition (e.g., Theory of Mind; Schilbach et al. 2010, 2013; Frith and Frith 2012), via neural reuse (Anderson 2010). There is empirical evidence suggesting that unconstrained cognition, emotional processing and social cognition might all share common neural networks in the dorso-medial prefrontal cortex and in the precuneus (Schilbach et al. 2012). Interestingly, the latter brain networks partially comprise the Default Mode Network, which is putatively activated more when a person does not directly focus on the outer world. Such a neural overlap between ‘social cognition’ and ‘introspection’ can be taken to suggest that not only thinking about others (either implicitly or explicitly), but even thinking about ourselves is driven by social interactions.

Taken together, we construe the self as a *historical process of dialectical attunement* unfolding over various time scales (Fig. 2). More concretely, we view two cardinal groups of processes dialectically interconnected, namely *internalization* and *externalization*. These processes are thought of unfolding along different time scales, e.g., (i) in the time frame of evolution, involving genetic and environmental adaptations, (ii) across generations, as cultural practices, or (iii) during individual development, including bodily and world reconfigurations, such as perception, action and learning. Put simply we view both low- and high-level attunement. Low-level attunement emerges during collective behaviour, when people are coupled together or when they

**Fig. 2** Dialectical attunement. Environmental structure (cf. social relations) is transformed within an individual via internalization processes (cf. predictive coding; rightward arrow). Internalized structures serve for co-regulating the external (social) world via externalization processes (cf. collective activity; leftward arrow). Internalization and externalization processes are thought of as unfolding dialectically, that is in a dynamic, reciprocal and cumulative interrelation. Please note, here schematic focus is put on the brain only for convenience; in reality the body participates in the dialectic of internalization/externalization as a whole



coordinate (cf. De Jaegher and Di Paolo 2007). However, while people interact, and thus act and perceive each other, they mutually co-construct internal models across multiple levels of bodily hierarchies. As we saw before, the construction of such hierarchies allows for consideration of increasingly higher levels of abstraction and thus temporal scales. That means that people in social interactions co-construct each other not only in the ‘here and now’, but also beyond, via co-configuring higher-level abstracted beliefs and patterns of action, on hand in future instances across a variety of interactive contexts or privately (cf. Theory of Mind; Fig. 2). Simply speaking, poetry (from the Greek “poiesis”, literally meaning “making”) can be thought of as an active externalization of internalized social interactions.

*Internalization* is the set of processes via which the structure of the environment (e.g., social relations) is actively transformed and implemented within an individual. From a Bayesian perspective, internalization entails the creation and maintenance of dynamic hierarchical models of the world in an effort to effectively predict future changes and act accordingly. We consider internalization as being accomplished across various time scales, from genetic information encoding and cultural adaptation, to bodily reconfiguration across development and real-time perception. For instance, in the evolutionary scale, the human visual system is attuned to the peak of the solar radiation spectrum that reaches the surface of the earth. In other words, human species has bodily internalized the environment in terms of electromagnetic conditions. Interestingly, similar attunement to environmental condition is also observed along developmental scales. For instance, experiments have demonstrated that extreme

exposure to a restricted range of visual stimuli (e.g., exclusively vertical visual orientation), early in development, modifies the morphology of neurons in visual cortex accordingly (e.g., Tieman and Hirsch 1982). Furthermore, with regards to shorter time scales, perception and action can be seen as real-time bodily attunement to the environment. Finally, undeniably people are also culturally attuned in multiple aspects. For instance, what is considered beautiful or delicious seems to be different across sociocultural contexts, both across time and space.

In fact, humans used cultural models for describing, predicting and manipulating the environment already in the cradles of civilization. For instance, ancient societies have construed natural phenomena, such as weather or earthquakes, as behavioural expressions of personified deities. At first sight, this might appear as a rather naive approach. However, we consider this as an ingenious tactic that might have allowed pre-scientific communities to recruit powerful cognitive capacities, originally developed for dealing with the undoubtedly complex social realm. Any level of abstraction can be considered as a model of the world. To come back to the example of language, a word can be thought of as a sociocultural model in and of itself, which of course presupposes the evolution of both the necessary biological apparatus across evolution and an interpersonal attunement across development. For instance, the word ‘animal’ or ‘wave’ practically captures and summarizes higher level similarities being met in a plethora of diverse natural processes. Here, we should stress that we do not consider the construction of internal models as a passive accumulation of representations.

The construction of internal models allows not only for the prediction of the world, but also the (socioculturally) transformation of it for meeting survival needs, through *collective externalization*. In other words, dialectical attunement does not merely imply a single-sided adjustment of the individual into the environment, but also transforming thereof across multiple scales: from cooking food, building shelters and developing technology, to transforming social structures and domesticating animal species. The activity of an individual in everyday life is decisively modulated by evolutionary, cultural and developmental factors. For instance, the use of a tool is defined by human anatomy, accumulated collective knowledge and individual learning. As discussed above, though, a change of the environment inherently entails a reconfiguration of the self as well. Externalization directly impacts on internalized models (cf. the interplay between active inference and predictive coding), as well as indirectly via the feedback of a transformed world. For example, learning to use a tool is fundamentally different when it is enacted rather than being merely theoretical, even though in both situations an internal model is developed. Additionally, both mechanical and conceptual tools (see the example of 'wave' from above) have helped the construction of modern technology, which in turn continuously modifies humans in multiple aspects and scales (from everyday behaviour to cultural habits and genetics in the long run). Crucially, when it comes to humans, transforming the world is fundamentally social, both with regards to our impact on others and the environment: the former is inherently social, while the latter becomes such via the mediation of sociocultural tools. In sum, we view the self exactly as the dialectic of the abovementioned internalization and externalization processes.

We will come back to this point and its scientific and societal implications during our concluding remarks (Sect. 2.3), after first describing how our hypotheses could be put to the test scientifically. To this end, we will describe experimental and data analytic means for studying the dialectic of internalization and externalization in *real-time* social interactions and beyond.

## 2.2 Two-Person Psychophysiology & Multi-level Accounts of Intersubjectivity

Due to conceptual and methodological constraints, research has largely focused on either intrapersonal (e.g. neurobiological and psychological), or interpersonal (e.g. sociocultural) processes. Here we emphasize the importance of studying intrapersonal and interpersonal processes in their inherent interrelation, as they unfold during social interactions. In what follows, we describe an experimental framework, namely two-person psychophysiology and an analysis

scheme, namely multi-level analysis of intersubjectivity that could help us do so.

*Two-person psychophysiology* appears as a promising avenue for empirical research, which while offering great experimental control, also preserves adequate degrees of ecological validity (Bolis and Schilbach 2017a, b). Traditionally, psychophysiology has enabled the empirical investigation of the relation between physiological and psychological processes (e.g., through physiological monitoring and introspection), offering important insights about individual mechanisms. However informative this kind of approach may have been, the concept of the (a-)typical 'self' will remain largely misconstrued until dynamic interpersonal processes are systematically considered, as social cognition might be fundamentally different when we are in interaction with others rather than merely observing them (Schilbach et al. 2013). It has been argued that the most important experience of the other comes from face-to-face situations; that this is the archetypic situation of social interaction, while all other situations are products of it (Berger and Luckmann 1967). It is exactly in this kind of situation that the 'here and now' of each other's subjectivity come together and possibly form an inextricable intersubjective unity (Berger and Luckmann 1967; De Jaegher and Di Paolo 2007; Bolis and Schilbach 2017b).

Building upon empirical frameworks of interpersonal research (e.g. Read Montague et al. 2001; Schilbach et al. 2006; Dumas et al. 2010; Barišić et al. 2013; Froese et al. 2015; Koike et al. 2016; Liu et al. 2016), two-person psychophysiology crucially allows for the empirical investigation and systematic manipulation of face-to-face social interaction, across various modalities and temporal scales. In such a framework (Bolis and Schilbach 2017b), participants sit opposite each other, working on tasks either individually or collectively, while being able to interact, either in real-time or offline, through a micro-camera communication system. Such a two-person framework allows for systematic control and monitoring of processes that live in different levels of description, from (epi-)genetics and culture to interpersonal behaviour and psychophysiology. In fact, via controlling the synchronicity of social interaction and composition of dyads, cardinal aspects of the self can be put into scientific test: *Emerging contextual and interpersonal differences in social interactions might prove equally, or even more important than individual traits in defining the becoming of the (a-)typical self* (Bolis et al. 2017).

Interpersonal frameworks for empirical research might be an important tool for moving beyond the individual as the unit of analysis, yet not sufficient on their own. Conceptual and experimental practices should be developed hand-in-hand with methods of analysis (e.g. Bahrami et al. 2010; Konvalinka and Roepstorff 2012; Schilbach et al. 2013; Abney et al. 2014; Dumas et al. 2014; Froese et al. 2015;

Friston and Frith 2015; Zapata-Fonseca et al. 2016; Fusaroli and Tylén 2016; Sevgi et al. 2016; Bolis and Schilbach 2017a). Here, we suggest a shift from an exclusive focus on the (Bayesian) brain in isolation, toward a *multilevel understanding of intersubjectivity and psychopathology*. In this framework of analysis, principled accounts of brain function (e.g. predictive processing) are employed for describing crucial neurobiological mechanisms, while being connected to real-life phenomena, which by definition live in an interpersonal space. More concretely, grounded in established models (e.g., Daunizeau et al. 2010; Mathys et al. 2011; Bolis et al. 2015), a two-level modelling scheme could be used for capturing both individual processes (Bayesian level) and collective behaviour (meta-Bayesian level). Put simply, in this scheme intrasubjective parameters will be deployed for capturing individual mechanisms (e.g., neuromodulation), while intersubjective ones to describe emergent processes on the collective level (e.g., interpersonal coupling). Collective parameters refer to sociocultural tools, such as artefacts, communication mediating factors, and generally any co-constructed and commonly held convention. For instance, the efficacy of a communication channel might strongly modulate interpersonal coupling in social interaction (Bolis and Schilbach 2017b).

Such an intersubjective scheme could be exploited for considering emergent phenomena on higher levels of description, such as for instance questions about the autonomy of a dyad or a group of people. To give a more specific example, in the context of collective externalization a non-linear model might explain observed behaviour optimally, thus, providing evidence that the group is different than the sum of individuals. Inversely, this framework could address questions about how collective processes, in turn, shape individual reality. For instance, one could differentially study the potentially distinct impact that a competitive or individualistic versus a collaborative structure might exert upon an individual (Bolis et al. 2017). Collective activity and societal structure are thought of being capable in shaping individual levels (from neurobiology to phenomenology) via internalizing mechanisms. In other words, it is not only lower-level mechanisms that result in emergent collective ones, but internal processes are treated, here, as dynamically internalized interpersonal processes.

Notably, a meta-Bayesian framework can consider observable activity in any level of description, such as neural activity, motor responses or collective behaviour. With regard to social interactions, an interesting avenue for future research might involve studying whether interpersonal coordination on the behavioural level might actually, serve as a prior and modulate, or even relax, the need for inferences about the hidden causes of social behaviour. Furthermore, at a neurobiological level, we hypothesize that activity of different neuromodulators could be related to a subject's ability

of tracking different levels of interpersonal regularities. In short, a Bayesian account of intersubjectivity intends to offer a principled and quantitative description of the dialectic between internalizing and externalizing processes across different levels of description, as discussed above.

### 2.3 The Dialectical Self: Scientific and Societal Relevance

Our approach shares common ground and most importantly brings together under a dialectical umbrella two seemingly disparate perspectives, i.e., interactionist-enactivist (e.g. Maturana and Varela 1980; De Jaegher and Di Paolo 2007) and computational-Bayesian accounts of cognition (e.g. Clark 2013; Friston 2013). Enactivist accounts have constructively put their focus on the fundamental role of interaction and coupling with the environment, including others. Bayesian accounts of cognition have provided important computational tools for describing individual cognition, mainly through hierarchical models. Our dialectical suggestion, on one hand emphasizes the primacy of (social) interactions. More concretely, it states that for a comprehensive understanding of the (a-)typical self, we will need to move beyond the individual, to the historical unfolding of (social) interactions over multiple scales. On the other hand, our approach extends Bayesian accounts of cognition by situating them in the context of real-time social interaction and providing a description of internalization and collective externalization processes beyond the individual. More precisely, it connects internalization to predictive coding and collective externalization to active inference. By doing so, it describes perception, learning and collective action as a unified process that allows for aligning personal (psycho-physiological) and interpersonal (coupling and synchrony) states with environmental (nature and others) conditions. Taken together, via integrating levels of description and time scales such an approach provides a unifying and principled way for studying the self beyond the individual.

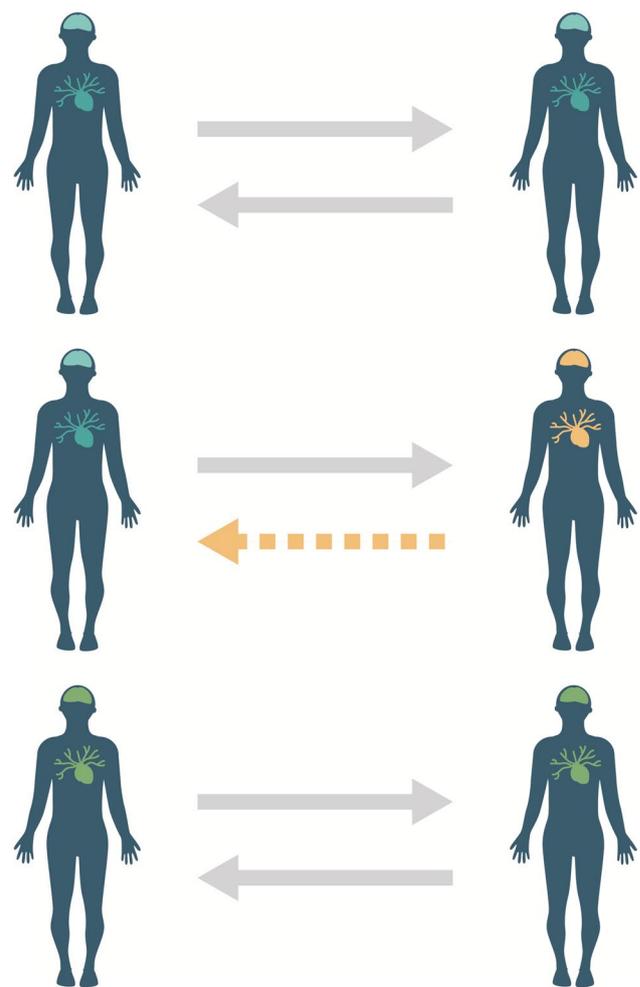
In this article we have described the self as the dialectic of internalization and externalization and more concretely as a historical product of dialectical attunement over various temporal scales (see Fig. 2). According to this view, low-level attunement is achieved largely automatically (beyond awareness) during embodied interactions, via mechanisms of collective externalization. High-level attunement is achieved through mechanisms of internalization. For instance, low-level attunement captures human action as an emergent collective phenomenon (cf. interpersonal bodily coupling, coordination and synergy) in the 'here and now'. High-level attunement captures human mind as an active environmental reflection. In a cultural frame, this takes the form of internalized values and conventions in a society, generalized across multiple temporal and contextual frames. In sum, low- and

high-level attunement are dynamically and cumulatively interrelated, via internalization and collective externalization processes, forming the *dialectical self*.

Yet still one might wonder why even question the question of the self. We believe that any thesis on the self is inherently implicated in numerous fields of science and the society. A dialectical perspective, as the one described here, points toward specific directions that acknowledge the primacy of the *social*, without neglecting the importance of the *individual* in their interrelation, co-construction and tension. Additionally, it points toward the necessity of adopting an empirical and principled approach to studying the self. To this end, formal approaches of predictive processing and dynamical systems appear as most promising. Approaching the formation of the self under the unifying umbrella of the dialectic of internalization/externalization might allow formal integration and re-description of seemingly disparate mechanisms across different scales. Yet the implications of such a dialectical approach reach further than the realm of scientific research.

In pedagogy, this is translated into an educational system that would promote collective problem solving as compared to mainstream competitive individual tests. Put simply, taking such an approach seriously, it would make no sense to isolate inherently limited individual cognitive capacity and reward merely the most relevant to a given task. On the contrary, promoting collective problem solving and decision making via active participation and interaction would enhance both cognitive and motivational aspects, yielding superior pedagogical but also practical achievements. In psychiatry, one would not be merely focused on diagnosing and 'fixing' individual impairments, but also tuning interpersonal communication and enhancing social inclusion (Fig. 3; Bolis et al. 2017). Within a clinical context, such an approach would suggest the monitoring of not only individual progress, but also interpersonal coupling between a 'therapist' and the 'individual', as well as between multiple persons during group therapy. In fact, not every therapist might be optimally suited for every patient and therefore matching of therapist and patient might need to be assessed in order to predict whether therapy will eventually work. Within a societal context, 'tuning' will not target only the individual with a psychiatric condition, but also her social environment. For instance, anti-stigma and informational campaigns will target tuning of social expectations of others as well, effectively resulting in a reciprocal amelioration of existing interpersonal misattunement. Such developments might help bring a redefinition of what a psychiatric disorder is, situating it back into the social realm within which it emerges.

In the field of ethics and law, seriously assimilating the idea that the self goes beyond the static individual, a juridical system would not only focus on individual intentionality



**Fig. 3** Dialectical (mis-)attunement and interpersonal re-tuning: (top) a homogeneous dyad interacting 'smoothly', (middle) a heterogeneous dyad interacting less effectively, (bottom) retuned interaction via not only targeting a person with a condition, but also others, as well as the interaction itself (cf. Bolis et al. 2017)

and responsibility, but also take into account collective factors and societal structure. Along similar lines, confronting social problems such as racism will not merely address educating individuals, but also dealing with social structures, which potentially instigate and maintain such patterns of behavior. Finally, such a perspective would suggest developing artificial intelligence and robotics, not via static pre-configuration, but via allowing interaction for co-constructing and internalizing knowledge. This should be expected to yield not only more robust artificial systems, but insightful conclusions on cardinal questions about human cognition as well. More concretely, in line with cultural historical and enactivist perspectives, we suggest that the role of social interaction and active participation in the co-construction of a culturally shaped self should be taken more seriously, in both research and social practice, as paraphrasing Descartes:

‘we interact, therefore I become’, or put simply ‘I interact, therefore I am’.

**Acknowledgements** Open access funding provided by Max Planck Society. DB would like to express his gratitude to colleagues in the research group for the developmental origins of human normativity (LMU Munich) for inspiring discussions during the 2015 reading group on Michael Tomasello’s work, Angeliki Angeletou for the beautiful artwork and Jessica Keverne for her diligent proofreading of this manuscript.

**Funding** This study was funded by a grant for an independent Max Planck Research Group to LS.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Research Involving Human and Animal Participants** This article does not contain any studies with human participants or animals performed by any of the authors.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Abney DH, Paxton A, Dale R, Kello CT (2014) Complexity matching in dyadic conversation. *J Exp Psychol General* 143(6):2304
- Anderson ML (2010) Neural reuse: a fundamental organizational principle of the brain. *Behav Brain Sci* 33:245–266. <https://doi.org/10.1017/S0140525X10000853>
- Bahrami B, Olsen K, Latham PE et al (2010) Optimally interacting minds. *Science* 329:1081–1085. <https://doi.org/10.1126/science.1185718>
- Barad K (2003) Posthumanist performativity: toward an understanding of how matter comes to matter. *Signs J Women Cult Soc* 28:801–831. <https://doi.org/10.1086/345321>
- Barišić I, Timmermans B, Pfeiffer UJ et al (2013) Using dual eye-tracking to investigate real time social interactions. CHI2013
- Baron-Cohen S (1991) Precursors to a theory of mind: understanding attention in others. *Nat Theor Mind* 1:233–251
- Barresi J, Martin R (2011) *History as prologue*. Oxford University Press, Oxford
- Berger PL, Luckmann T (1967) *The social construction of reality: a treatise in the sociology of knowledge*. Doubleday, New York
- Birhane A (2017) Descartes was wrong: “a person is a person through other persons” | Aeon Ideas. <https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other-persons>. Accessed 1 Nov 2017
- Blunden A (2000) Getting to know Hegel. Talk at Hegel, Marx and Derrida Summer School, University of Melbourne
- Bolis D, Becchio C, Schilbach L (2016) Revisiting psychological definitions at the interface of sociocultural historical theories and predictive coding. Welcome to WORLDING THE BRAIN: Patterns, Rhythms, Narratives in Neuroscience and the Humanities. p 10
- Bolis D, Schilbach L (2017a) Beyond one Bayesian brain: modeling intra- and inter-personal processes during social interaction: commentary on “mentalizing homeostasis: the social origins of interoceptive inference” by Fotopoulou & Tsakiris. *Neuropsychanalysis* 19:35–38. <https://doi.org/10.1080/15294145.2017.1295215>
- Bolis D, Schilbach L (2017b) Observing and participating in social interactions: action perception and action control across the autistic spectrum. *Dev Cogn Neurosci*. <https://doi.org/10.1016/j.dcn.2017.01.009>
- Bolis D, Heinzle J, Mathys C, Stephan KE (2015) Inferring cognitive traits of individual subjects through gaze controlled video games. In: Bernardis P, Fantoni C, Gerbino W (eds.) “TSPC2015: proceedings of the Trieste Symposium on Perception and Cognition, November 13rd 2015”, Trieste, EUT Edizioni Università di Trieste, 2015, p 19
- Bolis D, Balsters J, Wenderoth N et al (2017) Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. *Psychopathology*. <https://doi.org/10.1159/000484353>
- Brooks R, Meltzoff AN (2005) The development of gaze following and its relation to language. *Dev Sci* 8:535–543. <https://doi.org/10.1111/j.1467-7687.2005.00445.x>
- Bruner JS (1974) From communication to language—A psychological perspective. *Cognit* 3(3):255–287
- Carpenter M, Nagell K, Tomasello M et al (1998) Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr Soc Res Child Dev* 63:i. <https://doi.org/10.2307/1166214>
- Ciaunica A, Fotopoulou A (2017) The touched self: psychological and philosophical perspectives on proximal intersubjectivity and the self. In: Embodiment, enaction, and culture investigating the constitution of the shared world. MIT Press, New York, pp 173–192
- Clark A (2013) Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36:181–204. <https://doi.org/10.1017/S0140525X12000477>
- Dafermos M (2002) *The cultural historical theory of Vygotsky*. Atrapos, Athens
- Dafermos M (2015) Rethinking the crisis in social psychology: a dialectical perspective. *Soc Personal Psychol Compass* 9:394–405. <https://doi.org/10.1111/spc3.12187>
- Daunizeau J, den Ouden HEM, Pessiglione M et al (2010) Observing the observer (I): meta-Bayesian models of learning and decision-making. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0015554>
- De Jaegher H, Di Paolo E (2007) Participatory sense-making: an enactive approach to social cognition. *Phenomenol Cogn Sci* 6:485–507. <https://doi.org/10.1007/s11097-007-9076-9>
- Dobson S (2012) Face to face with the social brain. *Philos Trans R Soc Lond B* 367:1901–1908. <https://doi.org/10.1098/rstb.2011.0224>
- Dumas G, Nadel J, Soussignan R et al (2010) Inter-brain synchronization during social interaction. *PLoS ONE* 5:e12166. <https://doi.org/10.1371/journal.pone.0012166>
- Dumas G, de Guzman GC, Tognoli E, Kelso JAS (2014) The human dynamic clamp as a paradigm for social interaction. *Proc Natl Acad Sci* 111:E3726–E3734. <https://doi.org/10.1073/pnas.1407486111>
- Dunsworth HM, Warrener AG, Deacon T et al (2012) Metabolic hypothesis for human altriciality. *Proc Natl Acad Sci USA* 109:15212–15216. <https://doi.org/10.1073/pnas.1205282109>
- Ellison PT (2001) *On fertile ground*. Harvard University Press, Cambridge
- Fotopoulou A, Tsakiris M (2017) Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychanalysis* 19(1):3–28
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138. <https://doi.org/10.1038/nrn2787>

- Friston K (2013) Life as we know it. *J R Soc Interface* 10:20130475. <https://doi.org/10.1098/rsif.2013.0475>
- Friston K, Frith C (2015) A duet for one. *Conscious Cogn* 36:390–405
- Friston KJ, Daunizeau J, Kilner J, Kiebel SJ (2010) Action and behavior: a free-energy formulation. *Biol Cybern* 102:227–260. <https://doi.org/10.1007/s00422-010-0364-z>
- Frith CD, Frith U (2012) Mechanisms of social cognition. *Annu Rev Psychol* 63:287–313. <https://doi.org/10.1146/annurev-psych-120710-100449>
- Froese T, Iizuka H, Ikegami T (2015) Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Sci Rep* 4:3672. <https://doi.org/10.1038/srep03672>
- Fusaroli R, Tylén K (2016) Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. *Cogn Sci* 40:145–171. <https://doi.org/10.1111/cogs.12251>
- Grossmann I (2018) Dialecticism across the lifespan: towards a deeper understanding of ontogenetic and cultural origins of dialectical thinking and emotional experience. In: Spencer-Rogers J, Peng K (eds.) *The psychological and cultural foundations of East Asian cognition: contradiction, change, and holism*. Oxford University Press, Oxford
- Hobson RP, Patrick MPH, Crandell LE et al (2004) Maternal sensitivity and infant triadic communication. *Psychol Learn Motiv-Adv Res Theory* 45:470–480. <https://doi.org/10.1111/j.1469-7610.2004.00238.x>
- Hume D (1739/1888) *Treatise of Human Nature*. In: Selby-Bigge LA (ed). Clarendon Press, Oxford
- Jellema T, Baker CI, Wicker B, Perrett DI (2000) Neural representation for the perception of the intentionality of actions. *Brain Cogn* 44:280–302. <https://doi.org/10.1006/brcg.2000.1231>
- Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712–719. <https://doi.org/10.1016/j.tins.2004.10.007>
- Kobayashi H, Kohshima S (2001) Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J Hum Evol* 40(5):419–435
- Koike T, Tanabe HC, Okazaki S et al (2016) Neural substrates of shared attention as social memory: a hyperscanning functional magnetic resonance imaging study. *Neuroimage* 125:401–412. <https://doi.org/10.1016/j.neuroimage.2015.09.076>
- Konvalinka I, Roepstorff A (2012) The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front Hum Neurosci* 6:215. <https://doi.org/10.3389/fnhum.2012.00215>
- Kraut R (2017) Plato. In: Zalta EN (ed) *Stanford encyclopedia of philosophy*. Stanford University, Stanford
- Leontyev AN (1983) Activity, consciousness, and personality. In: Davydov V, Zinchenko V, Leontiev AA, Petrovskij A (eds) A.N. Leontiev. *Izbrannye psihologicheskie proizvedeniya* [A.N. Leontiev. Selected psychological works], vol 2. Pedagogika, Moscow, pp 94–231 (**Original work 1975**)
- Levins R, Lewontin RC (1985) *The dialectical biologist*. Harvard University Press, Cambridge
- Liu N, Mok C, Witt EE et al (2016) NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga Game with face-to-face communication. *Front Hum Neurosci* 10:82. <https://doi.org/10.3389/fnhum.2016.00082>
- Locke J (1690–4/1979) *An essay concerning human understanding*. In: Nidditch PH (ed) Oxford University Press, Oxford
- Magnus B (1970) *Heidegger's metahistory of philosophy: Amor Fati, being and truth*. Springer, Dordrecht
- Marx K, Engels F (1888) *The German ideology: including theses on Feuerbach and introduction to the critique of political economy*. Prometheus Books, New York
- Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011) A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5:39. <https://doi.org/10.3389/fnhum.2011.00039>
- Maturana HR, Varela FJ (1980) *Autopoiesis and cognition: the realization of the living*. Springer, Dordrecht
- Morales M, Mundy P, Delgado CEF et al (2000) Responding to joint attention across the 6-through 24-month age period and early language acquisition. *J Appl Dev Psychol* 21:283–298. [https://doi.org/10.1016/S0193-3973\(99\)00040-4](https://doi.org/10.1016/S0193-3973(99)00040-4)
- Mundy P, Newell L (2007) Attention, joint attention, and social cognition. *Curr Dir Psychol Sci* 16:269–274. <https://doi.org/10.1111/j.1467-8721.2007.00518.x>
- Mundy P, Block J, Delgado C et al (2007) Individual differences and the development of joint attention in infancy. *Child Dev* 78:938–954. <https://doi.org/10.1111/j.1467-8624.2007.01042.x>
- Nietzsche F (1999) *The birth of tragedy and other writings*. Cambridge University Press, Cambridge (**Original work 1872**)
- Nietzsche F (2017) *The will to power*. Penguin, Cambridge (**Original work 1901**)
- Powell JL, Lewis PA, Dunbar RIM et al (2010) Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia* 48:3554–3562. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2010.08.004>
- Ramstead MJD, Badcock PB, Friston KJ (2017) ScienceDirect Answering Schrödinger's question: a free-energy formulation. *Phys Life Rev* 11:541–546. <https://doi.org/10.1016/j.plrev.2017.09.001>
- Read Montague P, Berns GS, Cohen JD et al (2001) Commentary hyperscanning: simultaneous fMRI during linked social interactions. *NeuroImage*. <https://doi.org/10.1006/nimg.2002.1150>
- Roth W-M (2016) The primacy of the social and sociogenesis. *Integr Psychol Behav Sci* 50:122–141. <https://doi.org/10.1007/s12124-015-9331-5>
- Schilbach L (2016) Towards a second-person neuropsychiatry. *Philos Trans R Soc Lond B* 371:20150081. <https://doi.org/10.1098/rstb.2015.0081>
- Schilbach L, Wohlschlaeger AM, Kraemer NC et al (2006) Being with virtual others: neural correlates of social interaction. *Neuropsychologia* 44:718–730. <https://doi.org/10.1016/j.neuropsychologia.2005.07.017>
- Schilbach L, Wilms M, Eickhoff SB et al (2010) Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J Cogn Neurosci* 22:2702–2715. <https://doi.org/10.1162/jocn.2009.21401>
- Schilbach L, Bzdok D, Timmermans B, et al (2012) Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PLoS ONE* 7:e30920. <https://doi.org/10.1371/journal.pone.0030920>
- Schilbach L, Timmermans B, Reddy V et al (2013) Toward a second-person neuroscience. *Behav Brain Sci* 36:393–414. <https://doi.org/10.1017/S0140525X12000660>
- Seibt J (2017) Process philosophy. In: Zalta EN (ed) *Stanford encyclopedia of philosophy*. Stanford University, Stanford
- Sevgi M, Diaconescu AO, Tittgemeyer M, Schilbach L (2016) Social Bayes: using Bayesian modeling to study autistic trait-related differences in social cognition. *Biol Psychiatry* 80:112–119. <https://doi.org/10.1016/j.biopsych.2015.11.025>
- Sihvola J (2008) Aristotle on the individuality of self. In: Remes P, Sihvola J (eds) *Ancient philosophy of the self*. Springer, Dordrecht, pp 125–137
- Silbert LJ, Honey CJ, Simony E et al (2014) Coupled neural systems underlie the production and comprehension of naturalistic

- narrative speech. *Proc Natl Acad Sci USA* 111:E4687–E4696. <https://doi.org/10.1073/pnas.1323812111>
- Siller M, Sigman M (2002) The behaviors of parents of children with autism predict the subsequent development of their children's communication. *J Autism Dev Disord* 32:77–89. <https://doi.org/10.1023/A:1014884404276>
- Stetsenko A, Arievitch IM (2004) The self in cultural-historical activity theory reclaiming the unity of social and individual dimensions of human development. *Theory Psychol* 14:475–503. <https://doi.org/10.1177/0959354304044921>
- Striano T, Rochat P (1999) Developmental link between dyadic and triadic social competence in infancy. *Br J Dev Psychol* 17:551–562. <https://doi.org/10.1348/026151099165474>
- Tieman SB, Hirsch HVB (1982) Exposure to lines of only one orientation modifies dendritic morphology of cells in the visual cortex of the cat. *J Comp Neurol* 211:353–362. <https://doi.org/10.1002/cne.902110403>
- Timmermans B, Schilbach L, Pasquali A, Cleeremans A (2012) Higher order thoughts in action: consciousness as an unconscious re-description process. *Philos Trans R Soc Lond B* 367:1412–1423. <https://doi.org/10.1098/rstb.2011.0421>
- Tomasello M (1995) Joint attention as social cognition. In: Moore C, Dunham PJ (eds) *Joint attention: its origins and role in development*. Lawrence Erlbaum Associates, Hillsdale, pp 103–130
- Tomasello M (1999) *The cultural origins of human cognition*. Harvard University Press, Cambridge
- Tomasello M (2014) *A natural history of human thinking*. Harvard University Press, Cambridge
- Tomasello M, Carpenter M (2007) Shared intentionality. *Dev Sci* 10:121–125. <https://doi.org/10.1111/j.1467-7687.2007.00573.x>
- Tomasello M, Rakoczy H (2003) What makes human cognition unique? From individual to shared to collective intentionality. *Mind Lang* 18:121–147. <https://doi.org/10.1111/1468-0017.00217>
- Tomasello M, Carpenter M, Call J et al (2005) Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* 28:675–91–735. <https://doi.org/10.1017/S0140525X05000129>
- Vygotsky LS (1962) *Thought and language* (E. Hanfmann & G. Vakar, eds. and Trans). MIT Press, Cambridge, MA. **Original work 1934**. Translated in Greek (2008) by A. Rodi, Gnosi
- Vygotsky LS (1978) *Mind in society: the development of higher psychological processes*. Harvard University Press. **Original work 1930–1935**. Translated in Greek, (1997) by A. Bibou and S Vosniadou, Gutenberg
- Wong W (2006) understanding dialectical thinking from a cultural-historical perspective. *Philos Psychol* 19:239–260. <https://doi.org/10.1080/09515080500462420>
- Wong K (2012) Why humans give birth to helpless babies - Scientific American Blog Network. <https://blogs.scientificamerican.com/observations/why-humans-give-birth-to-helpless-babies/>. Accessed 1 Nov 2017
- Zapata-Fonseca L, Dotov D, Fossion R, Froese T (2016) Time-series analysis of embodied interaction: movement variability and complexity matching as dyadic properties. *Front Psychol* 7:1940. <https://doi.org/10.3389/fpsyg.2016.01940>