MICHELE DI FRANCESCO
Istituto Universitario di Studi Superiori, Pavia
michele.difrancesco@iusspavia.it

MASSIMO MARRAFFA
Università Roma Tre
massimo.marraffa@uniroma3.it

ALFREDO PATERNOSTER
Università di Bergamo
alfredo.paternoster@unibg.it

REAL SELVES? SUBJECTIVITY AND THE SUBPERSONAL MIND

abstract

The current philosophical discussion on the self and consciousness is characterized by a contrast or dilemma between the no-self (eliminativist) perspective, on the one hand, and the arguably naïve account that takes the self as a robust entity, on the other. In order to solve the dilemma, in this paper we suggest restoring a robust theory of the subject based on a bottom-up approach (fully consonant with contemporary neurocognitive science) together with a pluralistic reading of the nature of the science of the mental.

keywords

Self, bottom-up approach, explanatory pluralism, Dostoevskian Machine

1. Introduction and Overview

This paper was originally presented in a workshop addressing what was described as "Lynne Baker's Challenge", that is the thesis that human persons are entities essentially characterized by the possession of a *robust* first-person perspective (a thesis fully articulated by Baker in her recent book, *Naturalism and the First Person Perspective*, 2013). Differently from Baker's and many other talks presented in the workshop, the present contribution does not deal *directly* with the first-person perspective and its metaphysical implications. Rather it stems from the philosophical reflection on neurocognitive studies of subjectivity, and is more interested in *epistemological* and *explanatory* issues than in metaphysical conundrums. Yet it is fully congruent, we think, with Baker's appreciation of the importance of the relation between personal and subpersonal levels of explanation, as expressed, for example, in the following passage:

> Our ability to conceive ourselves as ourselves*[1] is a personal-level capacity. Why does it resist being reduced to or replaced by subpersonal phenomena? If I am right about the robust first-person perspective, then we have an answer to this methodological question: the personal level of reality – the level on which we live and love – is neither eliminable nor reducible to subpersonal levels that supply the mechanisms that make it possible for us to live and love (Baker 2014, p. 333).

We agree with Baker that the relation between personal and subpersonal "levels of reality" raises fundamental philosophical questions, and, among these, the problem of developing a theory of the nature of the self-conscious rational agent congruent with contemporary scientific research is one of the most prominent. We also take very seriously the "methodological" question addressed by Baker in the passage quoted above: why does our ability to conceive ourselves as ourselves* resist being reduced to or replaced by subpersonal phenomena? Indeed, in this paper we try to offer an answer to it; yet, differently from Baker's, our answer is based on a pluralistic reading of the nature of the science of the mental (which, as we shall see, involves a form of explanatory pluralism), rather than on a specific thesis about the metaphysical underpinnings of the first-person perspective.
In particular in our paper we argue that a *robust* account of the self – *i.e.*, of the subject of experience

1 The star following the second token of "ourselves" indicates a reference to the first person as a first person subject. You cannot substitute it salva veritate with a co-referential expression, such as the name of that person.

– is not only possible, *contra* the eliminativist-style arguments, but also fully consonant with contemporary (neuro)cognitive science. The paper is organized as follows.
In the first section we show that the contemporary science of the mind privileges a bottom-up approach to self-consciousness, based on the notion of cognitive, or computational, unconscious. In the second section we note that, in this context, the self-conscious rational agent is often presented as an illusion. A virtual space of presence, or a center of narrative gravity, is reconstructed as the owner of the stream of consciousness, but is in fact causally inert. In the third section we argue that a robust self is needed to explain the kind of intentional action and self-understanding presupposed by both commonsense psychology and social science. The problem we are faced to can then be presented in the form of a dilemma between the *no-self* (eliminativist) perspective, on the one hand, and the arguably *naïve* account that takes the self as a robust entity, on the other. In order to solve the dilemma, we suggest restoring a robust theory of the subject based on a bottom-up approach together with a pluralistic reading of the nature of the science of the mental. Also, we give some reasons to believe that this robust theory of the self is fully consonant with contemporary (neuro)cognitive science. Finally, in the fourth section, we compare our strategy with Baker's anti-eliminativist approach.
Before going on, we have to introduce a terminological *caveat*. For simplicity's sake, we use the word "self" in a loose way, to refer both to the subject of experience and to the self-representation of oneself that makes an individual a subject of experience. In other words, we do not use explicitly the distinction between "being a self" and "having a self". In fact we share Baker's doubts (or at least prudence) about the concept of self, and we consider this notion not as a primitive, but as a part of a theory of self-consciousness – which is the focus of our research.

2. The Cognitive Unconscious

In the last fifty years, the sciences of the mind have been mostly concerned with unconscious functions. Indeed, cognitive processes studied by cognitive science, such as perception, reasoning or language understanding, are not accessible to consciousness. Only their inputs and outputs (and perhaps some of their fragmentary parts) can be accessed. We are aware of the final results of the processes, not of their internal dynamics. In this perspective, the unconscious is in a way much more important than the conscious, insofar as it is the unconscious that explains the abilities manifested in our behavior.
Let us consider, for example, the case of language. Our understanding of a sentence is immediate. We instantly know whether or not we have grasped (as usually happens) what our interlocutor is telling us. Notwithstanding, a lot of machinery is needed to understand a sentence: a nearly continuous sequence of sounds must be segmented into words, *i.e.*, into meaningful units; a grammatical structure must be associated to the sentence, and not always this structure is the only possible one (in which case one needs to choose the right one); ambiguous or polysemous words are to be interpreted in a manner appropriate to the context, etc. We have no awareness of all these complicated processes, just as we have no awareness of the structures of information – the representations – that must be built up to successfully perform these tasks. We are not conscious of having grammar rules inside our heads and of systematically applying them during the processes of understanding.
The *cognitive* or *computational* unconscious, then, is a level of analysis that is fundamentally subpersonal: the information-processing level, wedged between the personal sphere of first-person phenomenology and the nonpersonal domain of neurobiological events. Such level no longer takes consciousness as something that explains, but rather as something that needs to be explained, analyzed, sometimes even dismantled.
In asking how consciousness, rather than the unconscious, is possible, the cognitive scientist fully endorses Darwin's methodological approach, which, assuming the continuity between animal and

human minds, pursues the study of consciousness by virtue of a bottom-up strategy. One begins with what is more simple, primitive, less structured, to reach what is complex, evolutionarily late, structured, without idealistically taking for granted the existence of a self-conscious self grounding the entire mental life. This self is rather the result of a process of construction that starts with subpersonal unconscious processes.

3. Dennett's Eliminativist Account of the Self

From the premise that the nature of the self is non-primary and derivative, many philosophers infer the conclusion that the self is an illusory by-product of the real neurobiological events, and is devoid of any explanatory role (think, for instance, of Dennett, Metzinger, analytical Buddhism). Let us focus on the case of Dennett, arguably the most influential one.
In light of a large amount of data from neurocognitive sciences, Dennett (1991) famously rejects the hypothesis that there is, in some area of the brain, a place where "it all comes together" (Dennett 1991, p. 107) – some sort of central executive system that coordinates all the cognitive operations – and stigmatizes it as "the myth of Cartesian Theater" (Dennett 1991, ch. 5). To this myth Dennett opposes the Multiple Drafts model of consciousness, according to which, at any instant, in any part of the brain, a multitude of "fixations of content" occur. The conscious character of these contents cannot be explained by their occurring in a *special* spatial or functional place (*i.e.*, the "Cartesian Theater"), nor by their having a special format. Rather, it depends on what Dennett (2005) calls "fame in the brain" or "cerebral celebrity" (Dennett 2005, p. 136). Like fame, consciousness is not an intrinsic property of the cerebral processes, but is more similar to "political clout", a kind of influence that determines the extent to which a content affects the future development of other contents distributed all over the brain.
On this eliminative view, a neuroscientific theory of consciousness must be a theory of how the illusion of the subject of consciousness arises (Dennett 1991, 2001, 2005). According to Dennett, an amazing property of *Homo Sapiens* is, precisely, the capacity to create a self: "out of its brain it spins a web of words and deeds" (1991, p. 416). By means of this activity, the biological organism produces a narrative, it posits a "center of narrative gravity". The narrative is the result of the working of a *Joycean Machine*: "In our brains there is a cobbled-together collection of specialist brain circuits, which, thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well-designed virtual machine" (1991, p. 228). The Joycean Machine is a *software in the brain* which creates the self, a *virtual captain*, a character described in internal and external discourse as the owner of the organism's mental states and as the actor of its actions and decisions, but who is in fact just a represented entity, not the real player in the game of human behavior.
Although Dennett's theory was developed in the early 1990s, recent empirical research is consonant with it. A neurocomputational architecture largely compatible with Dennett's Multiple Drafts Model is that of the Global Workspace Theory (GWT) of consciousness by Bernard Baars (1997). Recently, GWT has been developed in cognitive neuroscience, mainly thanks to Stanislas Dehaene and his collaborators' efforts (see, e.g., Dehaene & Naccache 2001; Dehaene *et al.* 2001). According to these researchers, there are two computational spaces within the brain, each characterized by a distinct pattern of connectivity.
The first space is a set of parallel, distributed, and functionally specialized processors or modular subsystems. These modular subsystems exploit highly specific local or medium-range connections that encapsulate information relevant to its function. The second space is a neuronal global workspace (and hence the theory is now termed "Global Neuronal Workspace Theory", GNWT) consisting of a distributed set of cortical neurons with long-distance connections, particularly dense in prefrontal, cingulate, and parietal regions, which are capable of interconnecting the multiple specialized processors and can broadcast signals at the brain scale in a spontaneous and sudden

manner. This global neuronal workspace breaks the modularity of the nervous system and allows the broadcasting of information to multiple neural targets. This broadcasting creates a global availability that is experienced as consciousness and results in reportability.
At least three features of the GNWT are significant for Dennett (see Schneider 2007, p. 318). First, it assumes that the neurocognitive architecture underlying the unity of consciousness is a distributed computational system with no central controller. Second, it makes massive use of recursive functional decomposition, an indispensable requirement to get rid of any homunculus who, nestled in a sort of incarnation of the pineal gland, scans the stream of consciousness. Third, it allows Dennett to hypothesize that the aforementioned "political clout" is achieved by "reverberation" in a "sustained amplification loop" of the winning contents (Dennett 2005, pp. 135-136).

4. The Dilemma: to Self or not to Self?

This eliminativist conclusion about the self is not necessitated by contemporary cognitive science; but, apparently, it is fully consonant with it. Cognitive (and neurocognitive) science starts from the idea of the fruitfulness of a bottom-up approach. This approach does not appeal to our introspective self-knowledge, but to the results of investigations into the gradual construction of human self-awareness: from the automatic and pre-reflexive construction of representations of the external world, through the bodily self-monitoring, to self-consciousness as introspective recognition of the presence of an *inner*, experiential space. The outcome is a criticism of the primacy of self-conscious subjectivity, where the latter, far from being a primary datum, becomes an articulate construction out of several neurocognitive and psychosocial components.
Another result of this approach is the acknowledgement of the mixed and multi-faceted nature of the self: minimal, autobiographical, narrative and social selves appear to reflect different aspects and different stages of the interaction between the neurocognitive and psychosocial components. Despite our appearance of unity, we are, in a literal sense, the product of the fusion of a wide range of composite processes.
We are then faced with a dilemma. Either we give up the classical notion of rational self-conscious agent; or we reject the eliminativist doubts about the self, and find a way to restore a robust theory of the self. We opt for the second horn of the dilemma, and in what follows we show how is possible to maintain a robust theory of the self, without sacrificing, at the same time, the merits of the bottom-up strategy. In other words, we stay with Dennett and neuroscientists in endorsing the bottom-up approach, but we stand apart from them in defending a robust account of the self.

4.1. A Strategy to Deal with the Dilemma

First of all, let us note that, in sketching a robust theory of the self, one should avoid the risk of falling again in anti-naturalist positions. Authors such as Alasdair MacIntyre, Charles Taylor, and Paul Ricoeur see the self as a self-interpreting being in a sense inspired by the hermeneutic tradition (Schechtman 2011). However, hermeneutic tradition is hardly compatible with the bottom-up approach, which involves a commitment to naturalism. A hermeneutical notion of self-interpretation, with its emphasis on meaning at the expense of the psychobiological theme of the unconscious, runs the risk of surreptitiously reintroducing the idealistic conception of the conscious subject as primary subject, since the subjectivity suggested by the hermeneuticists is inevitably intentionalizing – rather than intentionalized by – the unconscious. By contrast, we suggest that the self-interpretation is a theory-driven activity of narrative re-appropriation of the products of the neurocognitive unconscious, quite similar, for instance, to the notion developed by Peter Carruthers in his Interpretive Sensory-Access model of self-knowledge (see Carruthers 2011)[2].
To put it briefly, the robust theory of the self must not be a restatement of a top-down view of self-conscious subjectivity as a datum (the view of the subject as an *a priori*).

2 This does not mean that we buy Carruthers's theory of consciousness (and self-consciousness) across the board. The similarity concerns [just] the description of the activity of self-interpretation.

Somewhat surprisingly, Dennett sees narrativism and eliminativism about the self as the two sides of the same coin. In his view, the "I" is the useful fiction of a central controller, and its autobiography is a confabulatory by-product of the decentralized activity in the brain, which is actually responsible for the behavior. In other words, the Joycean Machine is anything but an *idle wheel* in the dynamical economy of the body (see Ismael 2006, p. 351). However, we dissent from the eliminativist argument that infers from the non-primary, derivative nature of the self a view of it as an epiphenomenal by-product of neurobiological events – or alternatively, of social (or socio-linguistic) practices – for at least two reasons. First, Dennett's self is said to be an abstraction devoid of real causal powers; yet, at the same time, this illusory character is a useful and even essential device: the complex social organisms that we are need a virtual self for their very survival, their social interactions, their decision making, and so on. Thus, one wonders why Dennett is not disposed to accord to the self a genuine causal role.
Second, let us concede that there is no central place in the brain where all information is gathered together, and no unifying superior function able to coordinate and organize what is processed by many different cognitive modules. This means that integration is not produced by top-down functions. However, this is not to say that the process of *ego* production leads to a pure nothing. We may argue on the contrary that the ability to represent herself as an enduring self does affect the very nature of the agent – making her an intentional subject of reason and action. To put it briefly, the inference from the existence of the *Multiple Drafts* to the *no-self* view is not justified.
Instead of Dennett's deflationist conception of the self, according to which the self is a mere abstraction, analogous to a non-existent, but useful, physical center of gravity, we suggest that there is an open alternative, a realist or somewhat *inflationist* position compatible with everything Dennett says about the architecture of human cognitive systems. According to such an inflationist option, the Joycean Machine is not a deceitful device but a cognitive mechanism that produces a reasonably stable and integrated (autobiographical) self, something that is best understood as the ongoing result of a narrative self-constructing process. Since the expression "Joycean Machine" may be perceived as intrinsically *eliminativistic*[3], we may substitute it with a new theoretical entity: the *Dostoevskian Machine*. The latter can be conceived as an integrated system of internal bottom-up mechanisms[4], which cooperate with external (social and environmental) factors to the process of self-building. A proper understanding of the working of the Dostoevskian Machine would reveal important aspects of human psychical dynamics, and would explain the processes that bring about the emergence of the kind of self-conscious experience that constitutes our autobiographical inner life and which shows itself, *inter alia*, in the use of self-referring linguistic expressions.
The reference to the Dostoevskian Machine allows us to save an important result of the eliminativist approach, namely, the acknowledgement of the mixed and multi-faceted nature of the self: minimal, autobiographical, narrative and social selves appear to reflect different aspects and different stages of the interaction between the neurocognitive and psychosocial components. Despite our appearance of unity, we are, in a literal sense, the product of the fusion of a wide range of composite processes.
To sum up, the bottom-up approach does not force us to endorse the eliminativist conclusions. In contemporary cognitive science there are theoretical tools which allow explaining conscious functions without assuming introspective self-knowledge as a datum, on the one hand, and maintaining a robust notion of the subject, on the other. We propose that self-conscious subjectivity, far from being a primary datum, is an articulate construction out of several neurocognitive and psychosocial components.
In other words, our central point is that the outcome of the Dostoevskian Machine, the product of the machinery in the head that composes the autobiography and controls verbal reports in the first person, *is responsible for stable, integrated and enduring aspects of human behavior.*

3 Thanks to Michael Pauen for this comment.
4 Here "bottom-up" means that these mechanisms are not based on any high-level, or full-blooded, representation of the self (this is in fact the *output* of the mechanism taken as a whole). Yet, some previous, relatively precocious, mental structures, such as bodily representations, feed the mechanism.

There are at least two considerations that can be invoked in favor of our robust-*cum*-naturalistic view of the self. Both have to do with (or partly involve) *hot* aspects of the mind. Let us explore them in turn.

4.2. Two Reasons for a (Naturalistic) Robust Theory of the Self

(*a*) The eliminativists disregard the fact that the process of narrative self-construction *includes an essential psychodynamic component*.
Breaking with a long philosophical tradition that has viewed self-consciousness as a purely cognitive phenomenon[5], the most important currents of dynamic psychology show that the construction of affectional bonds and the construction of identity cannot be separated. The description of the self that from 2-3 years of age the child feverishly 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). Briefly, the child needs a capacity to describe herself in a clear and consistent way, fully legitimized by the caregiver and socially valid. Also, this will continue to be the case during the entire cycle of life: the construction of an affectional life will always be intimately connected to the construction of a well-defined and interpersonally valid identity.
Accordingly, one cannot ascribe concreteness and solidity to one's own self-consciousness if the latter does not possess as a center a description of identity that must be clear and, indissolubly, "good" as worthy of being loved. Our mental balance rests on this feeling of solidly existing as an "I". If the self-description becomes uncertain (*i.e.*, inconsistent), the subject soon feels that her feeling of existing vanishes. This can be the result of a psychopathological process.
In patients with schizophrenia, for example, we can observe that the coherence of the representation of self is compromised or invalidated (see, e.g., Raffard *et al.* 2010), with a consequent loss of the capacity to clearly discriminate the borders between the inner space of the mind and the corporeal and extra-corporeal experiential spaces. The patient, then, develops abnormal defensive measures, aimed to head off the experiential chaos originating from the disintegration of the primary feeling of self.
Or let us consider the case of those patients whose main problem is a chronic feeling of insecurity (or lack of self-esteem, confidence in oneself, solidity of the *ego*, cohesion of the self – terms that we take to be essentially synonymous). According to a tradition in developmental psychopathology that begins with Michael Balint, Donald Winnicott and John Bowlby, the origin of this "basic fault" (Balint 1992) – or "primary ontological insecurity" (Laing 1960) – is to be traced back mainly to early deficiencies in the relationship between the child and the primary attachment figures (see, e.g., Fonagy, Gergely, Jurist & Target 2002). The child's attempts to rationalize the abusive or seriously neglective behaviors of the attachment figures may give rise to dysfunctional self-attributions, *i.e.*, to that *deficiency of identity* that can be found, for example, in patients suffering from narcissistic personality disorder. In some of these patients the feeling of identity is so precarious (the self is so little *cohesive*) that they find it difficult to feel existent and are afraid of completely losing contact with themselves if deprived of the link with situations, things or persons which serve as symbols that help to reassure them about their identity (Kohut 1977).
The waning of the existential feeling of presence may also occur in cases of sudden breakdown of self-esteem, or unexpected emotional upheavals, or when the continuity of the tissue of our sociality is broken, as can happen when one is suddenly thrown in some dehumanizing total institution (see, *e.g.*, the classic Goffman 1961). In such circumstances, the subject strives to cling to her memories, or to the sense of a projectual dignity, or to the secret security of an affiliation: "but if all these fail us, then we realize that our mind becomes empty, and not only we no longer know who we are, but also we literally lose the feeling of being present" (Jervis 2011, pp. 131-132).

5 See, e.g., Bermúdez: "Self-consciousness is primarily a cognitive, rather than an affective state" (2007, p. 456).

To recapitulate. The conception of self-consciousness that emerges from the bottom-up exploration of the mind – including a dynamic psychology driven by cognitive sciences – is that of an interminable process of self-objectification by the human organism. This consciousness of the self is a description of the self, namely, identity. In its most advanced form, this is finding oneself at the center of one's own orderly and meaningful subjective world, and hence at the center of a historical and cultural environment to which one feels to belong. However, this full-blown self-consciousness is a construction without metaphysical guarantee and thus it is not a faculty guaranteed once for all, being rather a precarious acquisition, continuously constructed by the human organism and constantly exposed to the risk of dissolution (see Marraffa 2013, p. 109).
This precariousness is the key to grasp the defensive nature of the Dostoevskian self-narrative. The construction and protection of an identity that is *valid* as far as possible is something rooted in the organism's primary need to subjectively subsist, and thus to solidly exist as "I". Thus, far from being an epiphenomenal, transient phenomenon, a character in a fiction invented to facilitate the prediction of behavior without any real correlate (a short-lasting *virtual captain*), the incessant construction and reconstruction of a cohesive self – *i.e.*, of an acceptable and adaptively functioning identity – is the process through which our intra- and inter-personal balances are produced, hence the foundation of our mental health. So, in contrast to Dennett's Joycean monologue, the Dostoevskian self-narrative is not empty chatter at all: it is a *causal* center of gravity.
On this view, the onset of self-consciousness is the establishment of a process of self-description, *i.e.*, the self-representing of a system encompassing mechanisms that interact across social, individual/personal, and subpersonal levels of organization (see Synofzik, Vosgerau & Newen 2008; Herschbach 2012; Thagard 2014). The description of identity imposes a teleology (focused on self-defense) on the system.

(*b*) The second consideration that can be invoked in favor of our robust-*cum*-naturalistic view of the self is grounded on the fact that the self-narration produced by the Dostoevskian Machine is not at all contingent and evanescent, since it is firmly anchored on *personality structures*.
Here we have in mind recent theoretical systematizations in personality psychology, where we find that the ability to perceive one's own identity in terms of *narrative identity* stems at least from two cognitive layers: (*i*) traits of personality, largely determined by genetic factors and substantially stable through the life cycle; (*ii*) goals, plans, projects, values and other constructs – *i.e.*, *motivational and strategic* roles and contexts – that define the life of an individual. Narrative identity is then an internalized and evolving story of the self – layered over the person's dispositional traits and characteristic goals and motives – which can provide the jumble of autobiographical memories "with some semblance of unity, purpose, and meaning" (McAdams & Olson 2011, p. 527).
Thus the experimental investigations on the mechanisms underlying the construction of identity can be seen as psychological hypotheses about the functioning of the extended or robust Dostoevskian Machine (rather than an evanescent and transient Joycean Machine). And here the reference to the necessity of a multi-level explanation of the robust Dostoevskian Machine comes in. These explanations, indeed, are located at the intersection of several psychological disciplines: personality psychology, social psychology, developmental psychology, dynamical psychology – all potentially interacting with neurocognitive research.
It is worth to point out that this involvement of a collection of different disciplines suggests, or even implies, a view of the explanatory practices in the sciences of the mind that can be dubbed as "explanatory pluralism". Let us say a few words on this.
Explanatory pluralism is a position in the philosophy of science holding that "theories at different levels of description, like psychology and neuroscience, can co-evolve, and mutually influence each other, without the higher-level theory being replaced by, or reduced to, the lower-level one"

(Looren de Jong 2001, p. 731). The need of increasing the available explanatory resources is the main concern of the pluralist, who distances himself both from the reductionist obsession for ontological parsimony and unification of science, and from the claim for strong autonomy of the special sciences theorists. In particular, against the reductionist claim that when lower-level explanations are completed, the higher-level explanations stop being causally explanatory, explanatory pluralists deny the existence of a *fundamental* explanatory level, and argue that higher-level entities continue to play a causal and explanatory role even when lower-level explanations are complete (see Marraffa & Paternoster 2013).

This is not the place for a detailed analysis of explanatory pluralism and its relevance to certain crucial, foundational issues in cognitive science (see, e.g., McCauley & Bechtel 2001; Craver 2007; Marraffa & Paternoster 2013). It is enough to point out that, as the considerations made in this section should have shown, the problem of giving a comprehensive explanation of self-consciousness, covering all its different levels, can only be addressed by means of a multiplicity of theoretical resources, stemming from different disciplines. In this sense, we take the issue of the self as a case for explanatory pluralism.

5. Concluding Remarks

We started our analysis referring to Lynne Baker's theory of the first-person perspective and to the connected claim that the personal level of reality is neither eliminable nor reducible to the subpersonal level (Baker 2013, 2014). So it could be useful ending with a comparison between Baker's defence of the irreducibility of the first-person perspective and our approach to self-consciousness. Firstly, we may note that there is a significant agreement on many issues. In particular:

- *We share a critical attitude towards reductive and eliminative accounts of mental phenomena, and in particular of the self.*
- *We share the idea that the personal level of description of human behavior, which characterizes the subject's mental life in terms of commonsense psychology, is essential and cannot be eliminated by direct reductive or eliminative moves.*

Besides, we believe that even if we grant a form of explicative supremacy of subpersonal psychology (an assumption that differentiates our position from Baker's), this does not entail the uselessness or the futility of personal psychology.

So there are convergences between the two theoretical projects, in particular if we consider our representation of the process of self-building as the product of what we called the "Dostoevskian Machine" (as opposed by Dennett's Joycean Machine).

We take the self as a system of subpersonal mechanisms that produces a real, causally efficacious, agent of psychical dynamics (and not a mere *virtual captain*), and this makes Baker's ontological view compatible with the kind of current empirical research we put at the basis of our proposal. This should not conceal the fact that Baker's overall metaphysics of the person is not the most favorable environment for a bottom-up approach to the subject, since it takes a person as a conscious substance (not an immaterial substance, but a substance which cannot be ontologically reduced to something else).

However, nothing in what we say forces *per se* a specific ontological conclusion. Non-reductive physicalism is compatible with the kind of explicative pluralism we endorse (in fact it is the standard view associated at the very beginning of cognitive science with explanatory pluralism). Yet, even some forms of metaphysical reductionism may be compatible: all depends on further and subtle issues concerning the metaphysics/epistemology divide.

If we were forced to express one ontological position that we find in accordance with our epistemological defence of the critical and fundamental role played by subpersonal explanation in psychology, we might quote David Lewis's seminal paper, "Attitudes *De Dicto* and *De Se*":

> I admit that knowledge *de dicto* is incomplete; but not that it is in any way misleading or distorted by its incompleteness. A map that is incomplete because the railways are left off is faulty indeed. By a misleading omission, it gives a distorted representation of the countryside. But if a map is made suitable for portable use by leaving off the "location of this map" dot, its incompleteness is not at all misleading. It cannot be said to misrepresent or distort the countryside at all, though indeed there is something that cannot be found out from it [...] An encyclopaedia that tells you where in logical space you are is none the worse for being neither signpost nor clock. Knowledge *de dicto* is not the whole of knowledge *de se*. But there is no contradiction, or conflict, or unbridgeable gap, or even tension, between knowledge *de dicto* and the rest. They fit together as nicely as you please (1979, p. 528; 1983, p. 144).

Adapting the quote to our analysis, we may say that the subpersonal description of human mind may be incomplete, but this does not mean that it is mistaken. We do not address this issue further, however.

We content ourselves with our attempt to show (1) that a more dialectical relationship between personal and subpersonal levels of psychological explanation is both possible and necessary to develop a theory of self-consciousness; (2) that a realist theory of the self offers an explanatory framework that is more useful to the understanding of self-consciousness than its eliminativistic anti-realist alternative; (3) that in the process of the construction of a theory of self-consciousness we need to wide our psychological horizon to take into consideration motivational and affective components that have been neglected by orthodox cognitive science, and (4) that this requires to widen our conceptual tools and suggests the adoption of epistemological pluralism[6].

6 We are grateful to the participants of the Workshop "Naturalism, the First Person Perspective and the Embodied Mind. Lynne Baker's Challenge: Metaphysical and Practical Approaches" (San Raffaele University, June 3rd–5th 2014), and in particular to Lynne Rudder Baker, Mario De Caro, Michael Pauen and Alfredo Tomasetta for their valuable comments.

REFERENCES

Baars, B. (1997), *In the Theater of Consciousness*, Oxford University Press, Oxford;
Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;
Baker, L.R. (2014), "The First Person Perspective and its Relation to Natural Science", in M. C. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?*, Routledge, London, pp. 318-333;
Balint, M. (1968/1992), *The Basic Fault: Therapeutic Aspects of Regression*, Northwestern University Press, Evanston (IL);
Bermúdez, J.L. (2007), "Self-Consciousness", in M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Blackwell, Oxford, pp. 456-467;
Carruthers, P. (2011), *The Opacity of Mind: The Cognitive Science of Self-Knowledge*, Oxford University Press, Oxford;
Craver, C.F. (2007), *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Clarendon Press, Oxford;
Dehaene, S. & Naccache, L. (2001), "Toward a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework", *Cognition*, 79, pp. 1-37;
Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.F., Poline, J.B. & Rivière, D. (2001), "Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming", *Nature Neuroscience*, 4(7), pp. 752-758;
Dennett, D. (1991), *Consciousness Explained*, Little Brown, Boston;
Dennett, D. (2001), "Are We Explaining Consciousness Yet?", *Cognition*, 79, pp. 221-237;
Dennett, D. (2005), *Sweet Dreams*, MIT Press, Cambridge (MA);
Fonagy, P., Gergely, G., Jurist, E.L. & Target, M. (2002), *Affect Regulation, Mentalization and the Development of the Self*, Other Press, New York;
Goffman, E. (1961), *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*, Doubleday, New York;
Herschbach, M. (2012), "On the Role of Social Interaction in Social Cognition: A Mechanistic Alternative to Enactivism", *Phenomenology and Cognitive Sciences*, 11, pp. 467-486;
Ismael, J. (2006), "Saving the Baby: Dennett on Autobiography, Agency, and the Self", *Philosophical Psychology*, 19(3), pp. 345-360;
Jervis, G. (2011), *Il mito dell'interiorità*, Bollati Boringhieri, Turin;
Kohut, H. (1977), *The Restoration of the Self*, International Universities Press, New York;
Laing, R. D. (1960), *The Divided Self: An Existential Study in Sanity and Madness*, Tavistock, London;
Lewis, D. (1979), "Attitudes *De Dicto* and *De Se*", *The Philosophical Review*, 88, pp. 513-543 (reprinted in Id. [1983], *Philosophical Papers*, Vol. I, Oxford University Press, Oxford, pp. 133-159);
Looren de Jong, H. (2001), "A Symposium on Explanatory Pluralism", *Theory & Psychology*, 11, pp. 731-735;
Marraffa, M. (2013), "De Martino, Jervis, and the Self-Defensive Nature of Self-Consciousness", *Paradigmi*, 31, pp. 109-124;
Marraffa, M. & Paternoster, A. (2013), "Functions, Levels and Mechanisms. Explanation in Cognitive Science and its Problems", *Theory & Psychology*, 1, pp. 22-45;
McAdams, D.P. & Olson, B.D. (2010), "Personality Development: Continuity and Change Over the Life Course", *Annual Review of Psychology*, 61, pp. 517-542;
McCauley, R.N. & Bechtel, W. (2001), "Explanatory Pluralism and the Heuristic Identity Theory", *Theory & Psychology*, 11, pp. 736-760;
Raffard, S., D'Argembeau, A., Lardi, C., Bayard, S., Boulenger, J.P. & Van der Linden, M. (2010), "Narrative Identity in Schizophrenia", *Consciousness and Cognition*, 19, pp. 328-340;
Schechtman, M. (2011), "The Narrative Self", in S. Gallagher (ed.), *The Oxford Handbook of the Self*, Oxford University Press, Oxford, pp. 394-416;

Schneider, S. (2007), "Daniel Dennett on the Nature of Consciousness", in M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Blackwell, Oxford, pp. 313-324;
Synofzik, M., Vosgerau, G. & Newen, A. (2008), "Beyond the Comparator Model: A Multifactorial Two-Steps Account of Agency", *Consciousness and Cognition*, 17(1), pp. 219-239;
Thagard, P. (2014), "The Self as a System of Multilevel Interacting Mechanisms", *Philosophical Psychology*, 27(2), pp. 145-163.