Dialogues in Philosophy, Mental and Neuro Sciences

Crossing Dialogues Association

ORIGINAL ARTICLE

The unconscious, consciousness, and the Self illusion

MICHELE DI FRANCESCO[a], MASSIMO MARRAFFA[b]

a: Faculty of Philosophy, University Vita-Salute S. Raffaele, Milan (Italy)
b: Department of Philosophy, University Roma Tre, Rome, Italy

In this article we explore the relationship between consciousness and the unconscious as it has taken shape within contemporary cognitive science — meaning by this term the mature cognitive science, which has fully incorporated the results of the neurosciences. In this framework we first compare the neurocognitive unconscious with the Freudian one, emphasizing the similarities and above all the differences between the two constructs. We then turn our attention to the implications of the centrality of unconscious processes in cognitive science for the classical conception of the self. Our analysis will bring to light a bit of claustrophobic dialectic between an eliminative variety of naturalism and an anti-naturalistic form of hermeneutics. Hence we conclude by recommending the pursuit of a mediation between such extreme stances.

Keywords: the unconscious, consciousness, intentionality, the self, cognitive sciences

DIAL PHIL MENT NEURO SCI 2013; 6(1): 10-22

THE PRIMACY OF CONSCIOUSNESS

Modern philosophy of mind originates with Descartes' identification of the mind with consciousness. This is a major paradigm shift from the Thomistic-Aristotelian tradition, and it could be useful to clarify its significance at the beginning of our analysis of the relation between conscious and unconscious mental states. In the second of his *Meditations on First Philosophy* Descartes defines the "thinking thing" as "[a] thing that doubts, understands, affirms, denies, is willing, is unwilling, and also that imagines and has sensory perceptions" (Descartes, 1641/1984, p.19). This list of the properties of the thinking thing groups together items that in the Aristotelian tradition were taken apart. In particular sensation and imagination are assimilated to intellective and conceptual parts of the mind. This assimilation, in turn, is due to a new way of understanding perception, which takes it as synonymous with "having a perceptual experience":

> For example, I am now seeing light, hearing a noise, feeling heat. But I am asleep, so all this is false. Yet I certainly *seem* to see, to hear, and to be warmed. This cannot be false; what is called 'having a sensory perception' is strictly just this, and in this restricted sense of the term is simply thinking. (Descartes, 1641/1984, p.19)

According to the Aristotelians, we need sense organs to perceive, and a perception requires bodily activity (and the existence of an 'external world'), while Descartes takes perception to be a conscious event confined within the boundaries of the conscious mind. As Richard Rorty has authoritatively argued, Descartes invents a sense of feeling as "no other than thinking", thus drifting away from the "Aristotelian distinction between reason-as-grasp-of-universals and the living body which takes care of sensation and motion". In doing so, Descartes makes room for a new distinction, that "between consciousness and what is not consciousness" (Rorty, 1979, p.51).

The result is both the identification of the mental with the conscious, and a peculiar view of the subject as a sort of inner observer of mental scenarios. Descartes creates a new conception of "the human mind as an inner space in which both pains and clear and distinct ideas passed in review before a single Inner Eye" (Rorty, 1979, p.50). We witness here the appearance in modern philosophy of what three centuries later Daniel Dennett would label the "Cartesian Theater" — which includes both the Inner Spectator and his inner mental world.

The centrality of consciousness was assumed also by critics of Descartes' philosophy, such as Locke, who bases his doctrine of personal identity on it:

> […] since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things, 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, 1690/1975, p.337)

We find in this short passage those relationships among consciousness, self, memory, responsibility for actions that characterize post-Cartesian modernity.

RESISTANCE TO THE IDEA OF AN UNCONSCIOUS MIND

The conception of the human mind as a unitary entity characterized by the primacy of consciousness was challenged first by Hume's skeptical philosophy, and then by Schopenhauer and Nietzsche. In the late 19th century, however, Cartesian mentalism was still prevailing, and indeed was exerting its influence on the early sciences of brain and mind. It is comprehensible, then, that philosophers, psychologists and neuroscientists were bewildered about phenomena (such as convulsive 'great' hysteria, dissociative fugue or multiple personality disorder) that appeared to be mental but went beyond the sphere of awareness and conscious control. They escaped the gaze of the Inner Eye.

As Livingstone Smith (1999) has convincingly shown, two strategies have been adopted to reconcile the existence of supposed unconscious mental phenomena with the consciousness-dependent conception of mind. The first option consisted in denying that such phenomena were genuinely *unconscious*; the evidence for unconscious mental states was reinterpreted as evidence for the possibility of a 'dissociation' or 'splitting' or 'doubling' of consciousness; namely, "the total possible consciousness may be split into parts which coexist but mutually ignore each other" (James, 1890/1950, p.206). The second option consisted in denying that such phenomena were genuinely *mental*; the evidence for the existence of unconscious mental states was reconceptualised as evidence for *neurophysiological dispositions* for genuinely (i.e. conscious) mental states.

These two strategies to bring apparently unconscious mental phenomena back to the traditional consciousness-dependent conception of the mind are still with us. Searle (1992), for example, has recast the dispositionalist approach to unconscious mental states, and in putting forward the claim that "the ontology of the unconscious is strictly the ontology of a neurophysiology capable of generating the conscious" has basically restated what the physiologist Ewald Hering had claimed in 1870 (cf. Livingstone Smith, 1999, p.141).

But the dissociationist strategy, too, has been revived by advocates of the so-called 'partitionist' approach to self-deception and its paradoxes. A first paradox is that it is logically incoherent and psychologically impossible that a single agent simultaneously holds contradictory beliefs. Another paradox is that since a successful interpersonal deception requires that A's intention remain hidden from B, a successful self-deception cannot occur, as in this case A and B are the same agent. The partitionist aims to dispel the first paradox by dividing the agent into two (or more) sub-agents, whose minds include the belief that *p* and the belief that non-*p* respectively, and tries to dissipate the second paradox by postulating that the deceived sub-agent cannot access the deceiving sub-agent's activities.

Donald Davidson is often considered the main 'partitioner', but his partitionism is actually very moderate. Davidson thinks that when one runs across what are traditionally seen as absurdities of Reason, such as akrasia or self-deception, the *personal* psychology framework is not to be given up in favor of the subpersonal one, but rather must be enlarged or extended so that the rationality set out by the principle of charity can be found elsewhere. On this perspective, the division of the mind is a *metaphoric* device employed to coherently describe (within the personal-level explanatory framework) a phenomenon (self-deception) that otherwise would be unintelligible. As Davidson puts it, a mental division is nothing but "a metaphorical wall" that keeps two contradictory beliefs separate. Consequently, we do not need to postulate "two minds each somehow able to act like an independent agent"; it is sufficient to imagine "a single mind not wholly integrated; a brain suffering from a perhaps temporary self-inflicted lobotomy" (Davidson, 1998, p.8).

A stronger version of partitionism was suggested by David Pears. Here the psychological partitioning is no longer Davidson's metaphorical wall; rather, it is a conceptual reconstruction of Freud's second topographical model of the mind. The psyche is divided into a "main system" and a "sub-system"; the latter is "built around the nucleus of the wish for the irrational belief" and is "organized like a person" (Pears 1984, p.87). Now, as Elster (1984) points out, Pears ascribes to the sub-system an internal rationality ("it is an efficient, quasi-altruistic manipulator of the main system"). This implies that the sub-system both has all sorts of propositional attitudes regarding the main system, and is "able to weigh and choose between alternative ways of satisfying the wishes of the main system". But then, Elster very properly concludes, "these requirements almost inexorably imply that the subsystem must have some kind of consciousness". In this connection, Laplanche and Pontalis (1967, entry "Topique") notice that Freud's second topographical model of the mind has an "anthropomorphic" character (cf. also Johnston, 1988, in which Pears' partitionism is labelled "homuncularist").

Thus we find here again that same need of reabsorbing the discourse on the unconscious into the discourse on consciousness that led some fin-de-siècle researchers to reinterpret the evidence for unconscious mental states as evidence for the possibility of a *dédoublement* of consciousness. On the basis of such a conclusion, it might appear strange that Davidson's (1982) and Pears' (1982) theories of self-deception are offered as *defenses* of Freud's theory. For is it not true that Freud put forward a subpersonal psychology (a 'metapsychology') that aimed to go beyond the psychology of consciousness? As a matter of fact, the psychological partitioning approach really captures an aspect of Freud's theory of the unconscious; unfortunately, however, it is an aspect that — as we will now see — represents a serious limitation of Freud's theory (cf. Marraffa, 2012).

THE UNCONSCIOUS: FROM FREUD TO NEUROCOGNITIVE SCIENCES

When, in the last decade of the 19th century, Freud intervened in the dispute on the unconscious, he first opted for a dispositionalist theory of (supposed) unconscious mental events. Then, in the context of the studies on hysteria conducted together with Breuer and strongly influenced by the French approach to neuropathology, Freud pronounced himself in favor of the dissociationist view. At least after the essay *The Neuro-Psychoses of Defence* (Freud, 1894/1962), however, the Viennese thinker tried to distance himself from the dispositionalist and dissociationist theories and to forge a theory of unconscious phenomena as occurrent and intrinsically unconscious mental events (cf. Livingstone-Smith, 1999, ch.5).

But there is a problem. Freud aims to go beyond the psychology of his times, which is a psychology of consciousness; his theory of the unconscious is, therefore, *programmatically* against psychological partitioning insofar as this treatment of self-deception remains — as we have seen — within a 'consciousness-centric' mentalistic framework. The problem is that, *as a matter of fact*, Freud failed to extricate himself from that framework. For, notwithstanding the revision of the Cartesian approach, in Freud the definition of the unconscious is still given *by its difference* from — and in some respects also *dependence upon* — the definition of consciousness; the latter is taken as a self-evident, primary datum, although it is then criticized and diminished in comparison with the traditional view (cf. Marraffa, 2012). And it is thus easy to notice that Freud tries to emancipate the sphere of the mental from consciousness only in "a few exceptional or anomalous cases (slips, neuroses etc.), and relative to a conception of mind as paradigmatically conscious" (Manson, 2000, p.163). And in those cases as well the Freudian unconscious is just an enlargement, or extension, of a psychology — folk psychology — hinged on the idea of a person who is able to have conscious mental experiences:

> [j]ust as much as the mental entities that parade across our consciousness, those that inhabit the [psychoana-

> lytic] unconscious are […] 'personal-level' phenomena […] in terms of their contents at least, unconscious ideas are conjectured to be indistinguishable from their conscious counterparts in all things save the fact that consciousness of them is absent. (O'Brien and Jureidini, 2003, p.143)

It may be remarked that in recent years a number of philosophers have argued that the extension of our ordinary psychological conception of mind is a strength of psychoanalytic theory. Here are two examples:

> [T]he grounds for psychoanalysis lie […] in its offering a unified explanation for phenomena (dreaming, psychopathology, mental conflict, sexuality, and so on) that commonsense psychology is unable, or poorly equipped, to explain. (Gardner, 1999, p.684)

> Freud […] proposed to add a deeper level of understanding than that provided by conscious psychology. (Nagel, 1994/1995, p.140)

This move may be taken as the basis of a defence of psychoanalysis against well-known epistemological challenges:

> Much of human mental life consists of complex events with multiple causes and background conditions that will never precisely recur. If we wish to understand real life, it is useless to demand repeatable experiments with strict controls. […] Explanations that refer to unconscious mental processes should be evaluated by the same standard. (Nagel, 1994/1995, p.142)

> Psychoanalytic explanations, like ordinary psychological explanations, may be exempt from the epistemological and methodological standards of experimental science. (Manson, 2003, p.179)

Once again Davidson is the referent of this conception of psychoanalysis. On his view, as is well known, the personal level is autonomous and different from the subpersonal, and is to be studied by means of different methods: you need hermeneutics, not the quest for natural laws. It is a view that has the quality of insisting on the unavoidability of the rational dimension of thought; and yet it is at risk of epistemological dualism — and avoids ontological dualism only by accepting the theory of the cause-reason relationship in *Mental Events* (Davidson, 1970). Although the issue is complex, Davidson's approach may be seen as a form of moderate antinaturalism that does not deny the physical unity of the world, but deprives science of the domain of the mental, which is construed as a space of reasons rather than causes, and is contrasted by the revisionary metaphysics of those philosophers whose reflection on neurocognitive sciences focuses on the following issue: how and to what extent should the folk-psychological conceptual framework be rectified in light of neurocognitive sciences? And it is clear that *if* we approach things in such a manner, the continuity with folk psychology is no longer a virtue of psychoanalytic theory. In this perspective, psychoanalysis moves from the personal to the sub-personal level, but then "it ends up having to re-import the personal level at the subpersonal, in order to get all the subpersonal bits to do what they are supposed to do" (Gardner, 2000, p.100). Briefly, psychoanalysis is a *personal* psychology that is masked as *subpersonal* psychology.

The response of the 'revisionary' philosopher to this difficulty of psychoanalysis will consist in opposing to the Freudian unconscious what cognitive scientists call the 'cognitive' or 'computational' unconscious. Here we find a level of analysis that aspires to be *genuinely* subpersonal: the information-processing level, wedged between the personal sphere of phenomenology and the subpersonal domain of neurobiological events. Such level no longer takes consciousness as an unquestionable assumption, as a non-negotiable given fact. The cognitivist mind is a process of construction and transformation of *representations*, where a mental representation is an explanatory hypothesis in a computational theory of cognition — a structure of information (somehow encoded in the brain), which is individuated exclusively in terms of its causal-functional role, and hence entirely apart from its (possible) phenomenological components. In brief, considerations concerning consciousness are not among the necessary and sufficient conditions that a mental state must satisfy to be qualified as representational. This methodological choice turned out to be extremely fruitful at an empirical level; the fact remains that — as we will mention later — it conceals within itself the seeds of some very thorny problems.

INTENTIONALITY AND CONSCIOUSNESS: WHAT COMES FIRST?

Neurocognitive sciences, therefore, challenge the traditional nexus between consciousness and intentionality, thus opening a conceptual space in which to build a theory of the 'non-derived' unconscious, viz. a theory that no longer obtains the unconscious by subtraction from consciousness. First one develops a theory of intentionality that is independent of and more fundamental than consciousness, a theory that makes no distinction between the various forms of unconscious representational mentality ("in brains, in computers, in evolution's 'recognition' of properties of selected designs" — Dennett, 1991, p.457). Then, one proceeds to work out a theory of consciousness on that foundation. In this perspective, consciousness is an advanced or derived mental phenomenon and not, as Descartes would have it, the foundation of everything mental. In short, first intentionality, then consciousness.

Viewing consciousness no longer as something that explains, but rather as something that needs to be explained, analyzed, dismantled, is also in full agreement with Darwinian naturalism. 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, i.e., reconstructing how the complex psychological functions underlying the adult self-conscious mind evolve from more basic ones. Again, if you like robust forms of naturalism, this is a clear advantage of the bottom-up, unconscious-centered strategy.

This bottom-up approach to human consciousness has been challenged by John Searle. For Searle's theory of consciousness leads us to consider illegitimate the hypothesis of any cognitive unconscious that cannot be accessible, even in principle, to consciousness. His main target is Noam Chomsky. As is well-known, the father of generative linguistics thinks that our capacity to produce/understand sentences rests on the knowledge of a universal grammar suitably parameterized and integrated by a lexicon. Chomsky characterizes this knowledge as unconscious or tacit. Now, there is no introspective route, even in the linguist's mind, that from the unconsciously 'cognized' principles of syntax leads to the awareness of them (cf. Rey, 1998). They are indeed unconscious contents, which cannot be introspected even in principle. Searle calls such contents "the deep unconscious", and against it he re-establishes an intrinsic link between intentionality and consciousness, the so-called "connection principle":

> Only a being that could have conscious intentional states could have intentional states at all, and every unconscious intentional state is at least potentially conscious. […T]here is a conceptual connection between consciousness and intentionality that has the consequence that a complete theory of intentionality requires an account of consciousness. (Searle, 1992, p.132)

Searle's connection principle implies the delegitimization of the Darwinian project of an understanding of the human mind's intentionality starting bottom up, i.e., from simpler and more basic intentional systems. The principle warns that the bottom-up strategy will never capture the level of human intentionality insofar as its essence is consciousness. And after undermining the possibility of viewing first-person experience as the presentation to consciousness of psychobiological functions of which we can reconstruct the ontogenetic and phylogenetic vicissitudes, Searle can only re-propose the Cartesian conception of consciousness as an entity that is primary, and grounds any other dimension of mental life.

This Cartesian conclusion brings Searle close to Brentano. On the one hand, Brentano puts forward a groundbreaking conception of consciousness as a set of forms of active relationship (i.e., construction of representations) between an organism and its environment-world. We are not conscious in the abstract, but are rather conscious of something; and among all the possible objects of the subject there is a particular object, the subject itself, whose representation is self-consciousness. Thus self-consciousness is not an entity that is to be idealistically conceived as a self-awareness that is primary, elemental, simple, preceding any other form of 'knowing'; rather it is a variation of our relationship to the world.

But Brentano takes two steps forward and one step back, for he holds that there is an intrinsic link between intentionality and consciousness that is very similar to that set up by Searle: "no mental phenomenon is possible without a correlative consciousness" (Brentano, 1874/1973, p.121). The same idea will be variously put forward by all the most important exponents of the phenomenological tradition. For example, according to Sartre the *conscience de soi* is "*the only mode of existence which is possible for a consciousness of something*" (Sartre, 1943/1956, p.20, italics in the text). And more recently, some philosophers have invoked Brentano's notion of "secondary consciousness", or Sartre's concept of "pre-reflective self-consciousness", in order to develop an alternative to the higher-order theories of consciousness. In the neo-Brentanian or neo-Sartrean perspective, any conscious experience is lived in a first-person perspective and conveys a primitive, pre-linguistic and preconceptual form of self-consciousness (cf. e.g. Gallagher and Zahavi, 2010). In spite of his proclaimed non-involvement in this tradition (Searle, 2008), Searle's claim fits in perfectly with it.

If we shift from these philosophers' a priori arguments to the psychological sciences, however, we find no evidence for such a primitive, prelinguistic and pre-conceptual form of self-consciousness. Rather, data from cognitive ethology and developmental psychology lead us to draw a *sharp* distinction between consciousness as a state of vigilance (being actively present to the world) and consciousness as self-consciousness (the agent's being present to herself). So we have empirical reasons to affirm that infants under one year of age are conscious in the sense that they are able to form a series of representations of objects and operational plans of action, and hence to interact with persons and things in flexible ways, but this occurs automatically, prereflexively, without any subjective experience of self-presentation, or cognition of a bodily or 'inner', experiential space (cf. Jervis, 2007, p.153).

Few species take a step beyond this basic interactive monitoring of the environment. Great apes like chimpanzees, and in our species infants from 15-18-20 months of age, can be said to attain a state in which they are able to make a clear distinction between their own physical bodies and the surrounding environment. More precisely, they first become capable of physical self-monitoring, i.e., focusing attention on the material agent as the (physical) executor of actions, and their bodily self-monitoring then comes to completion as the objectivation of a 'lived' body, and thus as a rudimentary self-consciousness.

It can be supposed that at an early stage human bodily self-consciousness, such as that of the chimpanzee, is structured by a nonverbal and analogic representation of the (physical) self, but very soon it begins to be mediated by the verbal exchange with the caregiver. In other words, in our species the chimpanzee-style, purely bodily self-consciousness is almost immediately outstripped and encompassed by a form of descriptive self-consciousness that is strictly linked to linguistic tools and social cognition mechanisms. Consequently, around the age of 3 or 4 years something occurs that can be observed only in the human species: the infant discovers that s/he has an 'inner life', i.e., s/he becomes able to identify and objectify his/her own subjectivity. Here the lived subjective experience takes as its object not only the outer world (as happens in all animals), not only the bodily world (as happens in chimpanzees and 15-18 month old children), but also *itself*: this is self-consciousness as *introspective* recognition of the presence of the virtual inner space of the mind, separated from the other two primary experiential spaces, viz. the corporeal and extracorporeal spaces.

Introspective awareness develops as *narrative identity* — i.e., by the end of the preschool years the child begins to experience herself as a person, to define herself as a certain kind of person, and to trace her own continuous identity as a person across time and space. This diachronic dimension of self-consciousness, viz. the possibility of tracing a unity that persists through time in our inner life, evolves as children attain a level of linguistic-narrative capacity which enables them to organize their own experiences in a chronological biography of self — a capacity that may not be fully consolidated until adolescence and early adulthood. This personal timeline that defines a continuous self through time is what Damasio terms the "autobiographical self", and which others define as "narrative identity".

By unearthing the non-primary but derived, constructed and partial character of self-consciousness, the cognitivist bottom-up approach can be regarded an *anti-phenomenology*, i.e., a critique of the subject, of its alleged givenness.

The term ‘anti-phenomenology’ was coined by Paul Ricoeur, who used it to define Freud’s methodological approach. According to Ricoeur, Freud’s establishment of the unconscious is “an *epochē* in reverse” because “what is initially best known, the conscious, is suspended and becomes the least known” (Ricoeur, 1965/1970, p.118). Consequently, whereas the phenomenological tradition pursues a reduction of phenomena *to* consciousness, capturing them as its objects, Freud’s methodological approach aims at a reduction *of* consciousness: the latter loses the Cartesian character of first and last certainty, which stops the chain of methodical doubts on the real, and becomes itself an object of doubt. However, as we have seen above, in reality Freud’s inquiry into the unconscious starts from consciousness taken as given, and this makes psychoanalysis more a dialectical variant of phenomenology than a complete dismissal of it. In contrast, sub-personal neurocognitive sciences, fortified by a consciousness-independent concept of intentionality, rightly qualify as an anti-phenomenology.

FROM HUME TO DENNETT (THROUGH FREUD): THE SELF ILLUSION

The preceding section allows us to estimate the distance that separates the new cognitivist mentalism from the “consciousness-centric” mentalism that characterized early experimental psychology, and from which the Freudian theory of the unconscious failed to disentangle itself. Under the influence of positivism, the introspectionist psychologists reified subjectivity. In most cases 19th century experimental psychology understood consciousness not in an experiential or subjective sense, but as an objective field, within which it was supposedly possible to break down mental contents, viewed as measurable objects. As an antidote to the positivistic attempt to reify phenomenological experience, the current sciences of mind and brain provide us with a repertoire of tools to penetrate the nature of introspective self-consciousness, making it possible to conceive phenomenological data not as tangible and measurable objects, but as the result of the presentation to consciousness of a collection of psychobiological functions, i.e., information-processing processes realized in biochemical events of the brain.

And what the neurocognitive-science toolkit allows us to claim about introspective self-consciousness is its fallibility (and, very often, its unreliability). Contrary to Cartesian and phenomenological insight, self-consciousness may be *deceptive* in its very nature. This is particularly true of the Spectator of the Cartesian Theatre: when the Inner Eye turns its gaze inward, the result is massive deception. As we will see below, a great number of data show that the description of the self as a description of identity is irreducibly out of phase, i.e. heterogeneous, with respect to the much more composite reality of the neurocognitive unconscious. This is not exactly a novelty. The problem had already been clearly framed by Freud, and before him by Hume.

Hume denies that we have experience of what we call our *self* (Descartes’ “myself” found in introspection; Locke’s person continuous and identical with herself across time; Damasio’s autobiographical self): “when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other […]. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception” (Hume,1739-40/1958, p.252). But then, what is the mind, if one can have experience of it only as a place of disparate perceptions? Hume’s answer is found in a famous passage:

> The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations. There is properly no *simplicity* in it at one time, nor *identity* in different; whatever natural propension we may have to imagine that simplicity and identity. The comparison of the theatre must not mislead us. They 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,1739-40/1958, p.253)

It is worth noticing that Hume is well aware that his conception of the illusory character of the unity of the mind owes us an explanation: “What then gives us so great a propension to ascribe an identity to these successive perceptions, and to suppose ourselves possest of an invariable and uninterrupted existence thro' the whole course of our lives?" (Hume,1739-40/1958, p.253). Hume's answer is grounded in a mixture of clever philosophical analysis (mainly on the notion of identity) and old-fashioned associationist psychology, whose details we needn't explore here. The result is the well known claim that the Self is just "a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (Hume,1739-40/1958, p.252). The unitary and continuous self is a fictional entity, perhaps a useful one, insofar as it gives our existence a sense of continuity, but metaphysically a fiction.

Freud's theory of person follows Hume's lesson and describes a *primary* self-deception when he sets up a contrast between the composite, non-monadical character of the mind and its unitary phenomenology. In the 1920s Freud develops his concept of Ego: the repertoire of automatic functions that convert the primitive energetic broth of the unconscious (Id) into thoughts, consciousness, responsible actions. In brief, the Ego is the very structure of the mind insofar as it is the organized part of what, when it is not structured, is the Id, and it is not entirely conscious. Now, in the "feeling of the Ego", Freud writes, the Ego "appears to us as something autonomous and unitary, marked off distinctly from everything else" (Freud, 1930/1961, pp.65-66). But it is a fallacious appearance: as a matter of fact the Ego is heterogeneous, heteronomous and secondary. Thus our mind is not self-transparent; on the contrary it eludes us, and also 'deceives' us; and it deceives us just starting from its pseudo-transparency and consciential pseudo-unity. The Ego owns non-truth-tropic cognitive mechanisms that generate the reassuring effect of a unitary egoic subjectivity that is master of the contents of consciousness. This effect is a "façade" whose deceptive character is to be denied if human beings are to feel their own autonomy, and thus experience themselves as persons.

Nowadays we can count on a large amount of behavioral, neuroimaging and computational investigations that offers robust evidence for both the claim that our mind's neurocomputational architecture is heterogeneous and decentralized, and the hypothesis that in presenting itself to consciousness such apparatus stages a complex self-deception. These two ideas — brightly prefigured by Hume and Freud — get a sophisticated philosophical and cognitive formulation in Dennett's theory of personal identity.

In light of a large amount of data from the neurocognitive sciences, Dennett (1991) rejects the hypothesis that there is, in some area of the brain, a place where "it all comes together" — some sort of central executive system that coordinates all the cognitive operations — and stigmatizes it as "the myth of Cartesian Theater". 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 referred to their occurring in a privileged spatial or functional place (i.e., the "Cartesian Theater"), and neither to their having a special format. It depends on what Dennett (2005) calls "fame in the brain" or "cerebral celebrity." Like "fame", consciousness is not an intrinsic property of the cerebral processes but is more similar to "political clout", i.e., the extent to which a content affects the future development of other contents distributed all over the brain.

A neurocomputational architecture that can be considered compatible with Dennett's Multiple Drafts Model is that of the Global Workspace Theory (GWT) of consciousness by Bernard Baars (1997). In this architecture consciousness is the global activation in a working memory — the global workspace in fact — whose contents can be broadcast to a wide range of cognitive systems. Recently, GWT has been developed in cognitive neuroscience, mainly thanks to Stanislas Dehaene and his collaborators' efforts (Dehaene and Naccache, 2001; Dehaene and Changeux, 2004). 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 (e.g., the elementary line segment detectors in area V1 or the motion processors in area MT, the 'visual word form' processor in the human fusiform gyrus, or the 'mirror-neuron' system in area F5). 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. 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 the umpteenth incarnation of the pineal gland, scans the stream of consciousness. Third, it allows him 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).

To sum up, the birth of a brain-based science of consciousness makes it possible to address in new ways the old question of the nature of the self, taking it from the philosopher's armchair to the laboratory of the cognitive scientist, so to speak. Dennett is just one of the growing number of scientists and philosophers who have been developing their attempts to solve Hume's problem. Our interest in his theory derives (i) from the consonance we may find between his analysis and Hume's (and Freud's), and (ii) from the substantial role played in his theorizing about the self by his view of the mind, conceived as a distributed computational system with no central controller.

It is worth noticing that, like Hume, Dennett thinks that, even if illusory, the appearance of the self must be explained; indeed the explanation of the illusion of the self is one of the main purposes of a theory of consciousness:

> A neuroscientific theory of consciousness must be a theory of the Subject of consciousness, one that analyzes this imagined central Executive into component parts, none of which can itself be a proper Subject. The apparent properties of consciousness that only make sense as *features enjoyed by the Subject* must thus also be decomposed and distributed … (Dennett, 2005, p.157)

Bottom-up, sub-personal, and third-personal approaches to the conscious mind cannot escape the difficult task of explaining how the higher-level, personal and first-personal perspective emerges; Dennett's eliminative view is that a neuroscientific theory of consciousness must be a theory of how the illusion of the Subject of consciousness arises. According to this philosopher, an amazing property of Homo Sapiens is, precisely, the capacity to create a self: "[o]ut of its brain it spins a web of words and deeds." (Dennett, 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. (Dennett, 1991, p.228)

The Joycean machine is this virtual machine, a "software in the brain" which, together with the organism and its cultural milieu, creates the self. Or better, it creates a "virtual captain," i.e., 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 in fact is just a represented entity, not the real player in the game of human behavior:

> Who's in charge? First one coalition and then another, shifting in ways that are not chaotic thanks to good

> meta-habits that tend to entrain coherent, purposeful sequences … (Dennett, 1991, p.228)

On Dennett's view, this inner character is just an abstraction, "not a thing in the brain." This seems to imply that a description of human action that invokes the self cannot be the ultimate truth. The real explanation, which involves real causes, will be found at brain level. In this sense a scientific theory of consciousness has to deal with the self not to postulate it as a real (causally efficacious) entity, but to banish the ghost of the first-person perspective from the neural machine.

NARRATIVISM, CONSCIOUSNESS, AND THE UNCONSCIOUS

Dennett's proposal can be seen as an eliminativistic version of the influential view of the self developed by the *narrativistic* tradition. Narrativism can be considered as one of the most important traditions of analysis of the self. More precisely, we may speak of the hermeneutical-narrative perspective, linking post-Heideggerian and post-Wittgensteinian thought with post-modern deconstructionism, to cover a truly far-reaching tradition that dominated 20th Century social science and philosophical anthropology (cf. Schechtman, 2011).

According to this perspective, "the self is constructed in and through narrative self-interpretations" (Gallagher and Zahavi, 2008, p.200). The self is not a (Cartesian) mental thing, nor a (Kantian) formal principle, but rather the result of a narrative *process*. The idea that the self is not the starting point of mental experience but the result of a *constructive process* is something that narrativism shares with the above-described bottom-up approach: the self is not a previously existing entity that produces the narrative, rather it is the product of the narrative itself.

But here we have to be cautious. For there is an anti-mentalistic version of this hermeneutical-narrative perspective which joins forces with social constructivism and linguistic idealism. From birth, each human being is involved in the process of self-creation, a process that requires constant input from society, and in which language is crucially involved. Here no margin is left for the 'information-processing' dimension of the agent: this sociolinguistic constructivism completely dismisses cognitive sciences, or tries to replace them with a 'psychology of the surface' which is relational and linguistic. So there are no information-processing mechanisms, not even mental states and processes: these things are opaque and unproductive; only relations and language hold. Thus this version of the hermeneutical-narrative perspective locates the self entirely within the public space of discourse. Harré (1993), for example, argued that psychological phenomena are produced in social interaction, and above all in the context of 'conversation,' beyond which there is no mental process; our conversational interactions *are* the mental processes (Harré, 1993). From here it is a short step to seeing persons not as the actors or the agents of discourses but rather as the products of the discursive practices themselves (cf. Harré, 1987). As Charles Taylor puts it, "to study persons is to study beings who only exist in, or are partly constituted by, a certain language" (Taylor, 1989, p.35).

Now, we definitely admit that the subject builds itself, *inter alia*, in social and conversational interactions, but being advocates of the bottom-up explanatory approach, we think that neither social structures nor linguistic and conversational schemes can be treated as Minerva born from the godhead of Jupiter with weapons. These things must be understood as *explananda*, and not as *explanantia*. For example, lexical acquisition invokes the mechanisms of mindreading; if children were not able to grasp the speaker's referential intentions, learning the meanings of words would not be possible (cf. Bloom, 2002).

But Dennett's naturalistic narrativism also has its problems. His approach definitely aims to be bottom-up and driven by cognitive sciences; nevertheless, this intent does not fit in well with Dennett's claim that the narrative self is a linguistic construction, where language is again idealistically taken as something given (cf. Cosentino, 2011). Furthermore, this philosopher sees narrativism and eliminativism about the self as two sides of the same coin; however, it is possible that narrativism does not entail the non-existence of the self, since it endorses the less radical claim that the self is created by a process of narration, and nothing prevents this created self from being causally efficacious. The Joycean machine metaphor does not imply that the created self is causally inert, only that the self is not a pre-existing entity — a position that few philosophers would nowadays endorse.

Indeed, classical narrativism, proposed within the hermeneutical tradition by scholars such as Alisdair MacIntyre, Charles Taylor, and Paul Ricoeur, does not take the self as a fiction. For example, according to MacIntyre narrative theory offers genuine explanations of human actions: to describe something as a human action we have "to identify it under a type of description which enables us to see that occurrence as flowing intelligibly from a human agent's intentions, motives, passions, and purposes" (MacIntyre, 1984, p.208; cf. also Schechtman, 2011, p.396). Here the self and her traits are essential components in an explanation that introduces "a normative or evaluative dimension" (Schechtman, 2011, p.396). This marks a major difference from Dennett's view:

> For Dennett the self is constituted through the human narration just as it is for the hermeneutical theorist, but there are important differences as well. Dennett's idea of narrative does not necessarily involve any strong form of evaluation or a quest for the good; it is more a matter of keeping track of the history of the body in which the narrating brain resides. […] In the former view [hermeneutical theory] there are genuine human selves, whose self-conception and mode-of-life constitute the selfhood: on the latter [Dennett's view] there are no such things. (Schechtman, 2011, p.396)

CONCLUDING REMARKS: HOW RADICAL MUST THE DECONSTRUCTION OF THE SUBJECT BE?

The points we have made in the preceding section show how a radical eliminative reading of the relationship between subpersonal and higher-level processes may not be the only one. The narrative construction of the self is an example; the neurobiological models of the self may be another. While discussing the relation between "the self and the issue of control", Antonio Damasio writes:

> Conscious deliberation, under the guidance of a robust self built on an organized autobiography and a defined identity, is a major consequence of consciousness, precisely the kind of achievement that gives the lie to the notion that consciousness is a useless epiphenomenon, a decoration without which brains would run the life-management business just as effectively and without the hassle. We cannot run our kind of life, in the physical and social environments that have become the human habitat, without reflective, conscious deliberation. But it is also the case that the products of conscious deliberation are significantly limited by a large array of nonconscious biases, some biologically set, some culturally acquired, and that the nonconscious control of action is also an issue to contend with. (Damasio, 2010, pp.271-272)

It is not necessary to go into further detail to notice that here we find a perspective quite different from the eliminative stance. But if this is so, and we allow reference to the self (and the personal perspective of the self-centered experience of the world) in the context of genuine explanations — if having a self makes a difference —, shouldn't we reconsider the quick dismissal of the conscious description of mental phenomena described in the first part of the paper?

The answer is yes and no. In a sense, nothing in what we have said makes the bottom-up strategy of the cognitive approach less valuable. The reading of the cognitivist approach that we have suggested here has the merit of emphasizing the theoretical import of the relationship of explanatory priority between conscious and unconscious phenomena: in a homuncularist, naturalist, and Darwinian perspective it is natural to expect that the former genetically and functionally depend on the latter. In the same context, one can hardly underestimate the importance of the criticism of the subject's conscious self-representation: if it is true that human beings "spin a self", as Dennett (1991, p.459) says, or more simply that each of us fabricates an inner narrative (a virtual, dynamical and interactive self-image: cf. Metzinger, 2009), there is a large amount of experimental evidence that challenges that image as a reliable account of the cognitive, affective, motivational processes that subserve our thoughts and behavior.

But we have also seen how a naturalist approach to the concepts of consciousness, subjectivity, and self does not necessarily require a complete deconstruction of the subject. By virtue of the creation of their 'egos' human beings have made a breakthrough, first biological and then socio-cultural, in navigating their world, and this beyond the limits of their capacities of self-representation. A view of the self as a result of the synthesis between biology and culture leaves open the possibility to assign a cognitive and explanatory (causal) role to some components of the folk-psychological discourse — avoiding both the creation of a radical rift between scientific psychology and social sciences, and the necessity of choosing between Dennett's eliminative naturalism and Davidson's anti-naturalist hermeneutics. Further developments of the debate might impose that choice on us, but at the moment the possibility of an intermediate point of view doesn't appear to be ruled out.

Corresponding Author:

Prof. Michele Di Francesco
Università Vita-Salute San Raffaele
Via Olgettina 58, 20132 Milano (Italy)
Phone: +39-022643 - 6178 / 5863
Fax +39-02643 6179
E-mail: difrancesco.michele@unisr.it

REFERENCES

Baars BJ. (1997) In the theater of consciousness: the workspace of the mind. Oxford University Press, Oxford.

Bloom P. (2002) How children learn the meaning of words. MIT Press, Cambridge MA.

Brentano F. (1874/1973) Psychology from an empirical standpoint. Routledge, London.

Cosentino E. (2011) Self in time and language. Conscious Cogn, 20:777-783.

Damasio A. (2010) Self comes to mind. constructing the conscious brain. Pantheon, New York.

Davidson D. (1970) Mental events. In: Foster L, Swanson JW. (Eds) Experience and theory. Duckworth, London:79-101.

Davidson D. (1982) Paradoxes of irrationality. In: Wollheim R, Hopkins J. (Eds) Philosophical essays on Freud. Cambridge University Press, Cambridge:289-305.

Davidson D. (1998) Who is fooled? In: Dupuy J. (Ed) Perspectives on self-deception. Cambridge University Press, Cambridge:1-19.

Dehaene S, Changeux JP. (2004) Neural mechanisms for access to consciousness. In Gazzaniga M. (Ed) The cognitive neurosciences III. MIT Press, Cambridge MA: 1145-1158.

Dehaene S, Naccache L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition, 79:1-37.

Dennett DC. (1991) Consciousness explained. Little, Brown and Company, New York.

Dennett DC. (2005) Sweet dreams. MIT Press, Cambridge MA.

Descartes R. (1641/1984) Meditations on first philosophy. In: Descartes R. The philosophical writings of Descartes, Volume II. Cambridge University Press, Cambridge:1-62.

Elster J. (1984) Managing to deceive ourselves. The Times Literary Supplement, 30 November.

Freud S. (1894/1962) The Neuro-Psychoses of Defence. In: Freud S. The standard edition of the complete psychological works of Sigmund Freud, Vol. 3. Hogarth Press and the Institute of Psychoanalysis, London:43-62.

Freud S. (1930/1961) Civilisation and its discontents. In: Freud S. The standard edition of the complete psychological works of Sigmund Freud, Vol. 21. Hogarth Press and the Institute of Psychoanalysis, London:57-146.

Gallagher S, Zahavi D. (2008) The phenomenological mind. Routledge, London.

Gallagher S, Zahavi D. (2010) Phenomenological approaches to self-consciousness. In Zalta EN. (Ed) The Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/archives/win2010/entries/self-consciousness-phenomenological/

Gardner S. (1999) Psychoanalysis, contemporary views. In: Wilson RA, Keil FC (Eds) The MIT Encyclopedia of the Cognitive Sciences. MIT Press, Cambridge MA:683-685.

Gardner S. (2000) Psychoanalysis and the personal/subpersonal distinction. Philos Explor, 3:96-119.

Harré R. (1993) Social being, II ed. Blackwell, Oxford.

Harré R. (1987) The social construction of selves. In: Yardley K, Honess T. (Eds) Self and identity: psychological perspectives. Wiley, New York:41-52.

Hume D. (1739-1740/1958) A Treatise of human nature. Oxford University Press, Oxford.

James W. (1890/1950) The principles of psychology. New York, Dover.

Jervis G. (2007) The unconscious. In: Marraffa M, De Caro M, Ferretti F. (Eds) Cartographies of the mind. Springer, Berlin:147-158.

Johnston M. (1988) Self-deception and the nature of mind. In: McLaughlin BB, Rorty A. (Eds) Perspectives on self-deception. University of California Press, Berkeley CA:63-91.

Laplanche J, Pontalis J-B. (1967) Vocabulaire de psychanalyse. PUF, Paris.

Livingstone Smith D. (1999) Freud's philosophy of the unconscious. Kluwer, Dordrecht.

Locke J. (1690/1975) An essay concerning human understanding. Clarendon Press, Oxford.

MacIntyre A. (1984) After virtue. University of Notre dame Press, Notre Dame Ind.

Manson N. (2000) A tumbling-ground for whimsies? The history and contemporary role of the conscious/unconscious contrast. In: Crane T, Patterson S. (Eds) The history of the mind-body problem. Routledge, London:148-168.

Manson N. (2003) Freud's own blend: functional analysis, idiographic explanation, and the extension of ordinary psychology. P Aristotelian Soc, 2:179-195.

Marraffa M. (2012) Remnants of psychoanalysis. Rethinking the psychodynamic approach to self-deception. Humana.Mente, 20:223-243.

Metzinger T. (2009) The Ego tunnel. Basic Books, New York.

Nagel T. (1994/1995) Freud's permanent revolution. In: Nagel T. Other minds. Critical essays 1969-1994. Oxford University Press, Oxford:26-44.

O'Brien G, Jureidini J. (2002) Dispensing with the dynamic unconscious. Philos Psychiatr Psychol, 9:141-153.

Pears D. (1982) Motivated irrationality, Freudian theory and cognitive dissonance. In Wollheim R, Hopkins J. (Eds) Philosophical essays on Freud. Cambridge University Press, Cambridge:264-288.

Pears D. (1984) Motivated irrationality. Oxford University Press, Oxford.

Rey G. (1998) Unconscious mental states. In: Craig E. (Ed) Routledge Encyclopedia of Philosophy. Routledge, London:522-527.

Ricoeur P. (1965/1970) Freud and philosophy: an essay on interpretation. Yale University Press, New Haven.

Rorty R. (1979) Philosophy and the mirror of nature. Princeton University Press, Princeton.

Sartre J.-P. (1943) Being and Nothingness. Philosophical Library, New York.

Schechtman M. (2011) The narrative self. In Gallagher S. (Ed) The Oxford Handbook of the Self. Oxford University Press, Oxford:394-416.

Searle J. (1992) The rediscovery of the mind. MIT Press, Cambridge MA.

Searle J. (2008) The phenomenological illusion. In: Searle J. Philosophy in a new century. Selected essays. Cambridge University Press, Cambridge:107-136.

Taylor C. (1989) Sources of the Self: The making of the modern identity. Harvard University Press, Cambridge MA.