

MEDICINA NEI SECOLI
ARTE E SCIENZA



GIORNALE DI STORIA DELLA MEDICINA
JOURNAL OF HISTORY OF MEDICINE

Fondato da / *Founded by* Luigi Stroppiana

QUADRIMESTRALE / *FOUR-MONTHLY*

NUOVA SERIE / *NEW SERIES*

VOL. 24 - No 1

ANNO / *YEAR* 2012

Articoli/Articles

*TROUBLES WITH SELF-CONSCIOUSNESS.
JERVIS ON INTROSPECTION AND DEFENSE MECHANISMS**

MASSIMO MARRAFFA

Department of Philosophy, University Roma Tre, I

SUMMARY

TROUBLES WITH SELF-CONSCIOUSNESS

Building on Sigmund Freud, Ernesto de Martino and cognitive sciences, Giovanni Jervis has outlined a view of introspective consciousness as primarily an activity of narrative re-appropriation of the outputs of the unconscious cognitive processing, emphasizing that such an activity is ruled by the primary need to construct an identity that is valid as much as possible. Thus Jervis has originally pursued an integration between the anti-introspectionist tradition in cognitive sciences and the psychodynamic investigation on defense mechanisms. This paper outlines Jervis' attempt to integrate these two traditions, and shows that it fits very well with recent developments of the theory-theory approach to self-knowledge.

Consciousness and self-consciousness

In his 1993 book *Fondamenti di psicologia dinamica* (Foundations of Dynamic Psychology), Jervis emphasizes the main shortcoming of the Freudian view of the unconscious¹. Freud's concept of the unconscious is derived *by subtraction* from the definition of consciousness. The latter is taken as a self-evident, primary quality of the mind, although it is then criticized and debunked relative to the idealistic conception of consciousness. Like all the psychoanalytic

Key words: Self-consciousness – Introspection - Defense mechanisms - Crisis of presence

ideas, the Freudian unconscious is a sort of enlargement or extension of our folk psychology:

Still today psychoanalytic theory is an intuitive theory of subjectivity (turned into objectivity) where subjectivity is intuitively extended “downwards”, i.e., to an unconscious world. The latter is “captured” and described with a sort of widespread literary good sense, which is rich in old and new reified fancies, such as “Ego”, “drives”, “censorship”, “internal nuclei”, “the phallus”, “Beta elements”, and so on almost endlessly².

Let us consider, by contrast, what cognitive scientists call the “cognitive” or “computational” unconscious. We find here a genuinely *subpersonal* level of analysis - the information-processing level, wedged between the personal sphere of phenomenology and the subpersonal domain of neurobiological facts - which no longer takes consciousness as an unquestionable assumption, as a non-negotiable given fact. The cognitivist mind consists in a process of construction and transformation of *representations*; and a mental representation is an explanatory hypothesis in a computational theory of cognition; it is a structure of information (somehow encoded in the brain), which is individuated exclusively in terms of intra-theoretical *functional* criteria.

In other words, cognitive sciences cut off the traditional nexus between consciousness and intentionality, thus opening a conceptual space to build a consciousness-independent conception of the unconscious. As Daniel Dennett famously puts it, first cognitive scientists develop a theory of intentionality that is independent of and more fundamental than consciousness - a theory that treats equally any form of unconscious representational mentality; and then, they proceed to build a theory of consciousness on that ground³.

In this perspective, one can easily realize that the ordinary term “consciousness” conflates two quite different kinds of psychological functions. First, consciousness is a state of vigilance, i.e., the agent’s

being actively present to the world: this is a matter of forming first-order representations of states of affairs, which may function to guide the agent's behavior. Second, consciousness is *self-consciousness*, i.e., the agent's being present to herself: this is a matter of being in a higher-order mental state, namely a representational state that has a first-order representation as its object. Studies in cognitive ethology and developmental psychology show that animals and infants under one year of age are conscious in the first-order sense of being conscious: they are able to automatically and pre-reflexively form a series of representations of states of affairs and operational plans of action, and hence to interact with persons and things in flexible *but not self-conscious* manners.

Only some species take a step beyond the basic interactive monitoring of the environment that characterizes the simple, primary consciousness of all animals. They attain self-consciousness, and this in at least two different senses.

Great apes like chimpanzees, and in our species infants from 15-18 months of age, can be said to reach 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 then their bodily self-monitoring comes to completion as the objectivation of one's own body (Merleau-Ponty's *corps propre*), and thus as a rudimentary self-consciousness.

However, it is only in human species, and only after the age of 3 or 4, that some psychological functions come to self-present themselves in compliance with the forms of self-conscious subjectivity. This is human consciousness in the full and traditional sense: self-consciousness as *introspective* recognition of the presence of the virtual inner space of the mind, separated from the other two primary existential spaces, i.e., the corporeal and extracorporeal spaces.

Experimental data show that the consciousness of the body as one's own physical body is necessary premise of the further formation of self-consciousness as consciousness of the existence of the mind as virtual inner dimension.

However, when this introspective consciousness is put under the magnifying lens of cognitive sciences, it is found to be something precarious, approximate, anything but guaranteed.

In the first place, Jervis notes, introspection is "a random, partial and unstable phenomenon" because "distractions, forgetfulness, partial awareness or temporary forgetting of learned knowledge constitute the vital fabric of our minds"⁴.

Furthermore, data from cultural psychology show that introspective consciousness is a sophisticated cognitive achievement, *incomplete* in many human beings. Both in children under 4-5 years, and in almost all normal adults in primitive pre-agricultural or *pre-literate* agricultural cultures, it can be observed the incompleteness of the capacity to conceptualize the existence of an inner space of the mind. This deficiency contributes to form a subjective psychological world that is conducive to somatic and pragmatic (rather than psychological) views of the agents: e.g., emotions are experienced as purely bodily feelings, and as such they are variously situated in guts or limbs; dreams are conceived as nocturnal visions located in the extracorporeal space; thinking is confused with speaking; and also one's own reveries and projects are only partially objectivated, and hence taken into consideration with difficulty. These forms of incompleteness of self-consciousness facilitate a proclivity to hysterical splitting and magic-religious rationalizations about themes of influencing, like jinx or possession⁵.

We can assume then that *our* self-consciousness - the self-consciousness of "us" as "educated" subjects who belong to an urban and industrial culture - is more *analytic* than that of the subjects who belong to pre-literate communities. However, this distinction between

the “primitive man” and the “civilized one” is much less clear-cut than it seems. For it is easy to note that even

*our assumption of responsibility for passions or moods that we ourselves [as “educated” subjects] produce is often incomplete. (With ease, sometimes also we say “I have been taken by...”, and believe to immediately obtain a sort of absolution from this.)*⁶

But the most important question that must be asked about *our* self-consciousness is the following:

*This re-appropriation of our inner life, which is the fulcrum of self-consciousness [...] as ground of an advanced ethics, and of which we are legitimately proud [...], well, is this more properly reflexive, introspective self-consciousness mainly “awareness”, or rather storytelling, or even conventional discourse?*⁷

According to Jervis, the answer that comes from some research traditions within cognitive sciences is that introspective consciousness is mainly a narrative construction that, as Freud saw it, “is saturated with self-deceptions and bad faith”⁸. This claim, and the empirical findings that lend support to it, are the topic of the next sections.

There is no introspective knowledge of the causes of our behavior and attitudes

Empirical evidence on the inaccuracy of introspective reports has been accumulating since the first observations about post-hypnotic suggestions in the late 19th century. Other landmarks in anti-introspectionism are Wilder Penfield’s experiments in stimulating the cortex, Roger Sperry’s and Michael Gazzaniga’s research on split brains, experiments on subliminal perception and dichotic listening, and above all the extensive cognitive dissonance and causal attribution literatures that have been built up in experimental social psychology over the last fifty years.

In R.E. Nisbett and T.D. Wilson's classical review of these literatures⁹, the experimental subjects' attitudes and behavior were caused by motivational factors inaccessible to consciousness. However, when explicitly asked about the *motivations* (causes) of their actions, the subjects did not hesitate to state their reasonable *motives*. The two psychologists explained this pattern of results by arguing that the participants did not have any direct access to the real causes of their attitudes and behavior; rather, they engaged in an activity of *confabulation*, i.e., they made use of *a priori causal theories* to build reasonable but imaginary explanations of the motivational factors of their attitudes and behavior¹⁰.

Nisbett and Wilson's account of causal self-attribution in terms of theory-laden confabulatory activity is an exemplar of "theory-theory account of self-knowledge"¹¹. For most theory-theorists, the attribution of psychological states to oneself (or first-person mindreading) is an interpretative activity that depends on mechanisms that exploit theories that apply to the same extent to ourselves and others. Such theory-driven mechanisms take as input information about mind-external states of affairs, essentially the target's behavior and/or the situation in which it occurs¹².

Evidence in support of the theory-theory account of self-knowledge comes from the so-called "actor-observer paradigm". In one application of this paradigm, Nisbett and Bellows compared the introspective reports of participants ("actors") to the reports of a control group of "observers" who were given a general description of the situation and asked to predict how the actors would react¹³. Observers' predictions were found to be statistically identical to -- and as inaccurate as -- the reports by the actors. This finding suggests that "both groups produced these reports via the same route, namely by applying or generating similar causal theories"¹⁴.

In developmental psychology Alison Gopnik has defended the theory-theory account of self-knowledge by arguing that there is good

developmental evidence of developmental *synchronies*: children's understanding of themselves proceeds in lockstep with their understanding of others¹⁵. E.g., since the theory-theory assumes that first-person and third-person mentalistic attributions are both subserved by the same theory of mind, it predicts that if the theory is not yet equipped to solve certain third-person false-belief problems, then the child should also be unable to perform the parallel first-person task. A much discussed instance of parallel performance on tasks for self and other is in a study by Gopnik and Astington¹⁶. Using the well-known "Smarties Box" paradigm, they found that the children's ability to answer the question concerning oneself was significantly correlated with their ability to answer the question concerning another. More recent meta-analytic findings show that performance on false-belief tasks for self and for others is virtually identical at all ages¹⁷.

Data from autism have also been used to motivate the claim that first-person and third-person mentalistic attribution have a common theoretical basis. An intensely debated piece of evidence comes from a study by Hurlburt, Happé and Frith, in which three people suffering from Asperger syndrome were tested with the descriptive experience sampling method¹⁸. The study showed marked qualitative differences in introspection in the autistic subjects: unlike normal subjects who report several different phenomenal state types (including inner verbalisation, visual images, unsymbolised thinking, and emotional feelings), the first two autistic subjects reported visual images only; the third subject could report no inner experience at all. Even though admitting that this is a very small sample of subjects, Carruthers suggested that "autistic subjects might have severe difficulties of access to their own occurrent thought processes and emotions"¹⁹.

Thus, data from social psychology, development psychology and cognitive neuropsychiatry converge on the idea that we have no introspection of the causes of our judgments, decisions and behavior.

In everyday life, explaining one's own motives ("being able to say why") plays mainly a *justificatory* role rather than a *descriptive* one²⁰:

Take, for example, the most basic of questions: "Why are you here?" If any individuals find themselves in a specific place at a certain time, it is unlikely they can pinpoint the interplay of factors or complex series of motives that have led them to be in that exact place at that precise time. But they will certainly have no hesitation in providing convincing explanations to justify their actions. In short, people can seldom say why they are there, but can always assert that it is right for them to be there²¹.

This allows us to unveil the illusory character of our coercive folk-psychological inclination to "read" any behavior as deliberately, consciously goal-directed, in accordance with a prior intention, which is supposed to be simple and identifiable²². Actually, Jervis argues, the agent is not a primarily *quiescent* organism, who "then" invariably moves towards some goal; rather it is a primarily *self-propelled* structure. Therefore, one can really say neither when one starts an action, nor when an identifiable goal-directed behavioral plan begins. It is more correct to say that

since ever we are embedded in a system of role scripts and behavioral schemes (i.e., of cognitive-motor schemes) which we have begun to articulate since when we exist as individuals, and which we restlessly modify and repropose according to the circumstances²³.

And embedded in this flow of actions, we sometimes tell ourselves: "This is just the thing I want to do", or "What I did is the thing that I really wanted to do", and again "This thought is just what I feel like thinking". In a word, we consider some piece of behavior as deliberate all the times that we are able to explain it with common sense, i.e., when we are able to justify it in accordance with the canons that are accepted in our social environment²⁴.

Thus introspection, insomuch as it is construed as a source of knowledge of the causes of our behavior and attitudes, is an illusion. In its

stead we find “the capacity to explain one’s actions *ex post*”, i.e., our continuously “praising” what we are doing²⁵.

There is no introspective knowledge of our attitudes

It is to be noticed, however, that the theory-theory account is never suggested as an *exhaustive* theory of self-knowledge; for some margin is always left for some sort of *direct* self-knowledge²⁶. Nisbett and Wilson, e.g., draw a sharp distinction between “cognitive processes” (the causal processes underlying judgments, decisions, emotions, sensations) and mental “content” (those judgments, decisions, emotions, sensations themselves)²⁷. Subjects have “direct access” to this mental content, and this allows them to know it “with near certainty”. In contrast, they have no access to the processes that cause behavior. However, insofar as Nisbett and Wilson do not offer any hypothesis about this alleged direct self-knowledge, their theory is incomplete.

In order to offer an account of this supposedly direct self-knowledge, some philosophers made a more or less radical return to various forms of Cartesianism, construing first-person mindreading as a process that permits the access to at least some mental phenomena in a relatively direct and non-interpretative way. On this perspective, introspective access does not appeal to theories that serve to interpret “external” information, but rather exploits mechanisms that can receive information about inner life through a relatively direct channel: this is the “inside access” view of introspection²⁸.

The inside access view comes in various forms. Mentalistic self-attribution may be realized by a mechanism that processes information about the functional profile of mental states, or their representational content, or both kinds of information²⁹. A representationalist-functional version of the inside access view is Shaun Nichols and Stephen Stich’s account of first-person mindreading in terms of “monitoring mechanisms”³⁰. Their hypothesis is that whereas detecting others’ mental states and reasoning about one’s own and oth-

ers' mental states are all subserved by the same theory of mind, the mechanism for detecting one's own mental states is quite independent of the mechanism that deals with the mental states of other people. More precisely, the monitoring-mechanism theory assumes the existence of a suite of distinct self-monitoring computational mechanisms, including one for monitoring and providing self-knowledge of one's own experiential states, and one for monitoring and providing self-knowledge of one's own propositional attitudes. Thus, e.g., if X believes that p , and the proper monitoring mechanism is activated, it copies the representation p in X 's Belief Box, embeds the copy in a representation schema of the form "I believe that___", and then places this second-order representation back in X 's Belief Box. Since the monitoring-mechanism theory assumes that first-person mindreading does not involve mechanisms of the sort that figure in third-person mindreading, it implies that the first capacity should be dissociable, both diachronically and synchronically, from the second. In support of this prediction Nichols and Stich cite developmental data to the effect that, on a wide range of tasks, instead of the parallel performance predicted by the theory-theory, children exhibit developmental *asynchronies*³¹. Moreover, they suggest³² that there is some evidence of a double dissociation between schizophrenic and autistic subjects: the monitoring mechanisms might be intact in autistics despite their impairment in third-person mindreading; in schizophrenics the pattern might be reversed.

The monitoring-mechanism theory provides a kind of Cartesian reply to the theory-theory. But it faces at least two difficulties. To start with, the theory must tell us how monitoring mechanisms establish which attitude type or percept type a given mental state belongs to³³. A possibility is that there is a separate monitoring mechanism for each propositional attitude type and for each perceptual modality. But then, since any monitoring mechanism can be selectively impaired, the monitoring-mechanism theory predicts a multitude

of dissociations – e.g., subjects who can self-attribute beliefs but not desires, or visual experiences but not auditory ones, and so on. However, the hypothesis of such a massive dissociability has little empirical plausibility³⁴.

A different version of the inside access view of introspection is Alvin Goldman's simulationist one. He argues that first-person mindreading both ontogenetically precedes and grounds third-person mindreading. Mindreaders need to introspectively access their offline products of mental simulation before they can project them onto the target. And this, Goldman claims, is a form of "direct access".

In 1993 Goldman put forward a *phenomenological* version of the inside access view, by arguing that introspection is a process of detection and classification of one's (current) psychological states that does not depend at all on theoretical knowledge, but rather occurs in virtue of information about the phenomenological properties of such states³⁵. But in light of heavy criticism³⁶, in his 2006 book Goldman has remarkably reappraised the relevance of the qualitative component for the detection of psychological states, pointing out the centrality of the *neural* properties³⁷. Building on Craig's account of interoception³⁸, as well as Marr's and Biederman's computational models of visual object recognition, Goldman now maintains that introspection is a perception-like process that involves a transduction mechanism that takes neural properties of mental states as input and outputs representations in a proprietary code - the introspective code, or the "I-code". The I-code represents types of mental categories and classifies mental-state tokens in terms of those categories. Thus Goldman's neural approach tries to solve the problem of the recognition of the percept and attitude types, which proved problematic for Nichols and Stich's representationalist-functionalist approach. However, since different percept and attitude types are presumably realized in different cerebral areas, each percept or attitude type will depend on a specific informational channel to feed the introspective mechanism. Consequently,

Goldman's theory also seems to be open to the objection of massive dissociability raised to the monitoring-mechanism theory³⁹.

Moreover, Peter Carruthers has offered a cogent argument against the idea of a direct access to propositional attitudes⁴⁰. His neuro-cognitive framework is the Global Neuronal Workspace model⁴¹, in which a range of perceptual systems "broadcast" their outputs (e.g., sensory data from the environment, imagery, somatosensory and proprioceptive data) to a complex of conceptual systems (judgment-forming, memory-forming, desire-forming, decision-making systems, and so forth). Among the conceptual systems there is also a multi-componential "mindreading system", which generates higher-order judgments about the mental states of others and of oneself. By virtue of receiving globally broadcast perceptual states as input, the mindreading system can easily recognize those percepts, generating self-attributions of the form "I see something red", "It hurts", and so on. But the mindreading system receives no input from the systems that generate propositional attitude events (like judging and deciding). Consequently, the mindreading system cannot directly self-attribute propositional attitude events; it must infer them by exploiting the perceptual input (together with the outputs of various memory systems). Thus, Carruthers concludes, "self-attributions of propositional attitude events like judging and deciding are always the result of a swift (and unconscious) process of self-interpretation"⁴².

On this perspective, therefore, we do not introspect our own propositional attitude events. Our only form of access to those events is via self-interpretation, turning our mindreading faculty upon ourselves and engaging in unconscious interpretation of our own behavior, physical circumstances, and sensory events like visual imagery and inner speech. Carruthers bases his proposal on considerations concerning the evolution of mindreading and metacognition, the rejection of the data that according to Nichols and Stich suggest developmental asynchronies and dissociation between self-attribution and

other-attribution, and above all on evidence about the confabulation of attitudes. (E.g., Carruthers discusses a study by Joaquim Brasil-Neto and colleagues⁴³, in which participants were caused to extend either their right or their left index finger via focal magnetic stimulation of areas of motor cortex in the relevant brain hemisphere. Participants had been instructed to make a free decision about which finger to extend after hearing a click, which was actually produced by the discharge of the magnetic coil. The participants confabulated: they claimed to have been aware of *deciding* to extend that finger.) Thus Carruthers develops a very sophisticated version of the *theory-theory* account of self-knowledge in which the theory-driven mechanisms underlying first- and third-person mindreading can count not only on observations of one's own behavior and the circumstances in which it occurs, but also on the recognition of a multitude of perceptual and quasi-perceptual events.

All this delivers us a drastically debunked conception of introspective consciousness. Although Carruthers admits that we can introspect conscious experiences like those that arise in perception and imagery, this perceptive and quasi-perceptive information is nothing but the raw material for an interpretative activity. In this perspective, then, the Cartesian soul-like, self-transparent consciousness-substance is superseded by the product of an apparatus -- our incomplete, partial, and in many cases seriously defective folk theory of psychology -- which allows us to very partially describe and above all narratively justify mental events and processes all fundamentally unconscious. Thus, like Hamlet's old mole, the unconscious has dug so deeply that, by this time, "it reigns over our whole mental existence"⁴⁴.

The self-defensive nature of self-consciousness

Another important hypothesis advanced by Carruthers is that Descartes' doctrine of the self-transparency of the mind reflects an innate feature of the human mind⁴⁵. That is, the mindreading system

would operate with a model of its own access to the rest of the mind that is essentially Cartesian, assuming that subjects know, immediately and without self-interpretation, what they are experiencing, judging and intending. This assumption, Carruthers speculates, (i) may have great heuristic value, greatly simplifying the mindreading system's computations; and (ii) may also "make it easier for subjects to engage in various kinds of adaptive self-deception, helping them build and maintain a solid self-image"⁴⁶.

This latter suggestion leads us to the topic of defense mechanisms. That is, we have now to focus on the *self-defensive* nature of our activity of re-appropriation of the outputs of the computational unconscious. Following Freud's lesson, Jervis holds that the construction of inner life is ruled by "a self-apologetic defensiveness"⁴⁷. But he adds the caveat that what we have learnt from cognitive sciences forces a radical revision of the way in which the psychoanalytic tradition has dealt with the study of defense mechanisms.

In his 1983 book *Presenza e identità* (Presence and Identity), Jervis distinguishes between two theories of error⁴⁸. The first theory is part and parcel with the Cartesian model of the mind. Descartes traced the errors of judgment and conduct back to the emotional, visceral, impulsive-instinctual, "animal" sphere of the body - this allowed him to safeguard the assumption of a primary (and for him transcendent) principle of human rational awareness. The Cartesian faith in reason as producer of truth, the idea that what is clear and distinct cannot be false, and that errors are essentially a sort of derailment due to drive-visceral interferences, is implicit also in Freud's system of thought. But the Cartesian conception of error had already found an implicit refutation in Francis Bacon's work, in which the errors of judgment and conduct are traced back to the forms of doing and knowing "that are peculiar to the psychological essence of human beings"⁴⁹. In Bacon, contrary to Descartes, the conscious and rational mind naturally produces errors: the human *intellectus*, he writes, "is like a false

mirror, which, receiving rays irregularly, distorts and discolours the nature of things by mingling its own nature with it”⁵⁰. We could say, in current terms, that Bacon sees the mind’s errors, illusions, and self-deceptions as intrinsic to the ordinary *cognitive-affective* processes. It is this Baconian perspective that has been taken by research traditions such as psychology of thought and social psychology. (Thus, e.g., social psychology tells us that stereotypes and prejudices are “structures of bad faith” that originate from cognitive mechanisms underlying the etiology of social attitudes⁵¹.) And in dynamic psychology the Baconian conception of error gives rise to “a reinforcing overturning” of the questioning about defenses. Now,

*the aspects of ambiguity, self-deception, and [...] sufferance of human life can no longer be conceived as interferences that are restrictively connected to affective and emotional factors (and hence negatively affecting a self-conscious rationality safeguarded as primary), but rather as globally constitutive aspects of the mind and behavior*⁵².

And thus,

*in examining the construction of the everyday life we need to explain not how and why some “defensive” mechanisms exist, but rather how all the structures of knowledge and action are by themselves, integrally, a matter of defenses*⁵³.

We are now able to capture something that is already in Freud but which the Cartesian model prevented him from thoroughly articulating: the defensive processes are something more than bulwarks against anxieties and insecurities that perturb the order of our inner life; actually, defense mechanisms are the very structure of the mind - the Freudian Ego itself is a defense. Here are the roots of the clinical theme of the fragility of the ego, namely that intimate personal insecurity that seems to originate from insufficiencies in the primary relationship between mother and child (what Michael Balint termed “basic fault”). But in Jervis the theme is much wider, and it concerns

the subject of a dynamic psychology based on the cognitive-science ontology of psychobiological functions. Let us see.

After undergoing the above-mentioned “reinforcing overturning”, the ideas of the unconscious and defense mechanisms have no longer the function of debunking the traditional image of a subject with a primary identity and force; on the contrary, they certify the non-existence of a human subject of that kind. What, more than anything else, defines the real human subject is its intrinsic fragility; and consequently what we must try to understand is how the human subject, notwithstanding its fragility, is able to *construct* itself:

Today’s scientific psychology wants to understand how the human subject comes to exist and survive, and to be conscious, and to create culture, despite the fragility of the biological premises from which it starts. It may be said that the problem is no longer to know how human beings can “come down” from the level of nobility at which they were placed, but on the contrary, how they can “rise” up to self-consciousness and culture in spite of the lack of an identity and a strength that guarantee them; notwithstanding, therefore, their ontological insubstantiality, and indeed, still more radically, a sort of their original “non-being”⁵⁴.

Jervis’ emphasis on the precarious nature of the subject’s self-construction reveals the influence of his mentor Ernesto de Martino⁵⁵. Central in de Martino’s work is the concept of *presence*, viz. the finding oneself again at the center of a one’s own orderly and meaningful subjective world, and hence at the center of an historical and cultural environment to which one feels to belong. Presence includes, however, the opposition in the form of the risk of its dissolution, i.e. of the pure negativity, of the fragmentation of the elements outside of the synthesis of manifold. As de Martino writes in his 1948 book *Il mondo magico* (The Magic World):

[even] the supreme principle of the transcendental unity of self-consciousness involves a supreme risk to the person, i.e., the risk of losing the supreme principle through which it is constituted and established⁵⁶.

According to the great philosopher-ethnologist, therefore, the human presence in the world is a precarious acquisition, continuously constructed by the subject and constantly exposed to the risk of the crisis (the “crisis of presence”). Jervis elegantly describes this idea:

*the feeling of existing, that is, the primary feeling of the presence of self to itself, or if you like the feeling of the unity of the ego, or also the self-consciousness as full certainty on which the experience and the order of everyday living rest, is not a psychological faculty guaranteed once for all, but it is a precarious acquisition, arduously constructed by culture every day*⁵⁷.

The exclusive focus on culture is what separates de Martino from Jervis.

De Martino thought that overcoming the crisis of presence is prerogative of culture; indeed, presence is cultural dynamism:

*Presence is movement that transcends the situation in value. Owing to this movement, it is detached from the situation, emerges from it, moulds it as a situation in an “operable” world. Presence is man’s primordial ethos, the will of history that unceasingly unfolds*⁵⁸.

If presence is movement, the crisis is inactivity, a state of stand-still. The “critical moments of becoming” are just those situations in which the inertia of presence, which is equivalent to its loss, becomes an imminent threat. This may occur in the confrontation with death, in cases of psychological dissociation, alienation, and loss of subjectivity⁵⁹. In the resolution of these moments the religious *dehisterification* -- in which becoming is replaced by the “iteration of the identical” -- plays a fundamental role⁶⁰.

All these themes undergo a naturalistic reversal in Jervis’ writings. He tries to distill de Martino’s psychological insights from his historicist and culturalist thought. Thus, de Martino becomes a forerunner of the current sociological and psychological research on identity and the self⁶¹.

The problem of identity is the main aspect of de Martino's thought that is of interest to psychology. At the pages 46-50 of the aforementioned *Presenza e identità*, Jervis opposes de Martino's theory of presence to Kant's conception of self-consciousness. Kant thought that the consciousness of existing could be captured in a pure state, independently from the consciousness of existing in a certain way: "I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am", he famously writes in the first *Critique*⁶². But de Martino's presence cannot be conceived as a *primum*, a self-awareness that is primary, elemental and simple, preceding any other form of "knowing". As Jervis puts it,

*the being-here as the subject's primary feeling, open to the world, gets its meaning only as a being here in a certain way, i.e., as bodily and affective self-image, in representing to itself one's own person as a person of a certain type. There is no feeling of self without some form of representation of self: contrary to what Kant thought, there is no consciousness of self without knowledge of self*⁶³.

In other words, we know that we exist insofar as we know that we exist "in a certain way", as *describable identity*, constant through changes. If for some reason this self-description becomes uncertain, the subject soon feels that the feeling of existing vanishes⁶⁴.

Self-consciousness is, therefore, the construction of a description of self. And grafting de Martino's "phenomenological psychology of identity" onto dynamic psychology, Jervis identifies the ultimate root of the primary defensiveness of the self-constructed subject in the precariousness of this description of identity - "the primary, and universal, existential risk of the *loss of presence*"⁶⁵. Without the ontological guarantee on which the Cartesian soul-like consciousness-substance can count, the human subject constitutes itself as a repertoire of "composite psychological manoeuvres"⁶⁶, which take pains to cope with its "ontological insubstantiality", its "original non-being". We have here a hypothesis about human nature:

Troubles with self - consciousness

Our life does not respond only to those elementary biological needs, surviving and reproducing, which are manifest to everyone; and neither our inner drives can be traced back only to the universal forms of social competition, which we see in the rivalry among animals. By contrast, our everyday life is also conditioned by a need that, in the human species at least, is as much fundamental, namely the need to construct and protect a self-image endowed with at least a minimal solidity, and that is, in practice, solid enough to confirm to ourselves that we exist without dissolving ourselves⁶⁷.

In Jervis' hands, then, de Martino's inquiry into the universal structures that allow individuals defend themselves from anxiety in view of the critical moments of becoming, is naturalized and incorporated in a dynamic psychology that investigates the uncertainties that concern self-consciousness by means of cognitive-science tools⁶⁸. And thus Jervis' dynamic psychology deals with the thematic of the presence and its crisis as a matter certainly historical but also, and perhaps still more, biological and psychological. Accordingly, unlike de Martino, Jervis places the defense mechanisms along a continuum that spans all levels, from the individual to the collective. At the individual level, the defensive activities consist in "the intrapsychic defenses and the interpersonal manoeuvres to which each of us appeals, in the relationship with other people and one's own environment, to defend one's own self-describability and, indissolubly, the cohesiveness of one's own self-conscious being"⁶⁹. At the collective level, such activities consist in the construction of "a system of references, in part symbolic and ritual, which give perspective to living, domesticity and meaning to one's own being in the world"⁷⁰.

BIBLIOGRAPHY AND NOTES

- * This essay is one of a series of five papers in which I have been trying to reconstruct and develop Jervis' work on three themes: the unconscious, consciousness, and identity. The other four essays are: Jervis e la genealogia nascosta della coscienza umana. In: JERVIS G., *Il mito dell'interiorità. Tra psicologia e filosofia*, edited by G. Corbellini and M. Marraffa. Turin, Bollati Boringhieri, 2011, pp. XI-LVI; JERVIS G., De Martino e il mito dell'interiorità. *Rivista di Filosofia* 2011; 2: 241-59; *Precariousness and Bad Faith. Jervis on the Illusions of Self-Conscious Subjectivity*. *Iris. European Journal of Philosophy and Public Debate* 2011; 3(6): 171-87; *Remnants of Psychoanalysis. Rethinking the Psychodynamic Approach to Self-Deception*. *Humana.Mente - A Journal of Philosophical Studies* 2012; 20: 149-50.
1. JERVIS G., *Fondamenti di psicologia dinamica*. Milano, Feltrinelli, 1993, pp. 200-1.
 2. JERVIS G., *A proposito dell'editoriale di Paolo Migone del n. 2/2004*. *Psicoterapia e scienze umane* 2004; 38, 4: 499. The translations of Jervis' writings are all mine except for the quotations from the essay quoted in note 4.
 3. DENNETT D.C., *Consciousness Explained*. New York, Little, Brown & Co., 1991, p. 457.
 4. JERVIS G., *The Unconscious*. In: MARRAFFA M., DE CARO M. e FERRETTI F. (eds.), *Cartographies of the Mind*. Berlin, Springer, 2007, p. 155.
 5. At the end of 1950s Jervis investigated these phenomena in the context of a team study of the ecstatic healing cult of tarantism in the Salentine Peninsula of southern Italy. See: DE MARTINO E., *La terra del rimorso*. Milan, Il Saggiatore, 1961; Engl. trans. *The Land of Remorse*. London, Free Association Books, 2005.
 6. JERVIS G., *Sopravvivere al millennio*. Milan, Garzanti, 1995, pp. 74-75.
 7. JERVIS G., *Il mito dell'interiorità. Tra psicologia e filosofia*. Turin, Bollati Boringhieri, 2011, p. 195.
 8. JERVIS G., see note 7, p. 88. (JERVIS G., note 4, pp. 149-50) assesses Freud's hypothesis of a systematic tendency towards self-deception within our everyday thought processes as the *strength* of his theory of the unconscious.
 9. NISBETT R.E. and WILSON T.D., *Telling more than we can know: Verbal reports on mental processes*. *Psychological Review* 1977; 84: 231-59.
 10. See also JOHANSSON P., HALL L., SIKSTRÖM S., OLSSON A., *Failure to detect mismatches between intention and outcome in a simple decision task*. *Science* 2005; 310: 116-19, where Nisbett and Wilson's legacy is

developed through a new experimental paradigm to study introspection, the “choice blindness” paradigm.

11. This is a variant of a *symmetrical* or *self/other parity* account of self-knowledge: see SCHWITZGEBEL E., *Introspection*. In: ZALTA E.N. (ed.), *The Stanford Encyclopedia of Philosophy*. URL=<<http://plato.stanford.edu/archives/fall2010/entries/introspection/>>, §2.1. See also ROBBINS P., *The ins and outs of introspection*. *Philosophy Compass* 2006; 1(6): 619, who calls such account “outside access view of introspection”.
12. The main historical referent of such an approach to self-knowledge is Daryl Bem’s “self-perception theory”. See BEM D.J., *Self-Perception Theory*. In: BERKOWITZ L., (editor), *Advances in Experimental Social Psychology*. New York, Academic Press, 1972, vol. 6, pp. 1-62.
13. NISBETT R.E. and BELLOWS N., *Verbal reports about causal influences on social judgments: Private access versus public theories*. *Journal of Personality and Social Psychology* 1977; 35: 613-24.
14. NISBETT R.E. and WILSON T.D., see note 9, pp. 250-1; see also SCHWITZGEBEL E., see note 11, §§2.1.2 and 4.2.1.
15. GOPNIK A., *How we read our own minds: The illusion of first-person knowledge of intentionality*. *Behavioral and Brain Sciences* 1993; 16: 1-14.
16. GOPNIK A. and ASTINGTON J.W., *Children’s understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction*. *Child Development* 1988; 59: 26-37.
17. WELLMAN H.M., CROSS D. and WATSON J., *Meta-analysis of theory-of-mind development: The truth about false belief*. *Child Development* 2001; 72: 655-84.
18. HURLBURT R., HAPPÉ F. and FRITH U., *Sampling the form of inner experience in three adults with Asperger syndrome*. *Psychological Medicine* 1994; 24: 385-95. In this experimental paradigm, subjects are instructed to carry a random beeper, pay attention to the experience that was ongoing at the moment of the beep, and jot down notes about that now-immediately-past experience. See HURLBURT R.T. and SCHWITZGEBEL E., *Describing Inner Experience? Proponent Meets Skeptic*. Cambridge (MA), MIT Press, 2007.
19. CARRUTHERS P., *Autism as mind-blindness: An elaboration and partial defense*. In: CARRUTHERS P. and SMITH P.K. (eds.), *Theories of Theories of Mind*. Cambridge, Cambridge University Press, 1996, p. 261. See also FRITH U. and HAPPÉ F., *Theory of mind and self-consciousness: What is it like to be autistic?* *Mind & Language* 1999; 14(1): 14. More recent evidence

is in WILLIAMS D. and HAPPÉ F., “What did I say?” versus “What did I think?”: *Attributing false beliefs to self amongst children with and without autism*. *Journal of Autism and Developmental Disorders* 2009; 39: 865-73; Idd., *Representing intentions in self and other: Studies of autism and typical development*. *Developmental Science* 2010; 13: 307-19.

20. In this sense, such research trends link themselves to a tradition of critical thought that refers not only to Freud’s notion of rationalization but also Marx’s concept of ideology.
21. JERVIS G., see note 4, pp. 156-7.
22. JERVIS G., see note 1, p. 184.
23. JERVIS G., see note 1, p. 257.
24. JERVIS G., *Prime lezioni di psicologia*. Roma-Bari, Laterza, 1999, p. 90.
25. JERVIS G., note 4, p. 157.
26. SCHWITZGEBEL E., see note 11, §2.1.3.
27. NISBETT R. and WILSON T., see note 9, p. 255.
28. ROBBINS P., see note 11, p. 618. “Self-detection account of self-knowledge” in SCHWITZGEBEL E., see note 11, §2.2.
29. ROBBINS P., see note 11, p. 618.
30. NICHOLS S. and STICH S., *Mindreading*. Oxford, Oxford University Press, 2003.
31. E.g., children are capable of attributing knowledge and ignorance to themselves before they are capable of attributing those states to others. See WIMMER H., HOGREFE G. and PERNER J., *Children’s understanding of informational access as a source of knowledge*. *Child Development* 1988; 59: 386-96.
32. On the basis, inter alia, of a reinterpretation of Hurlburt, Happé and Frith’s data, quoted in note 18.
33. GOLDMAN A.I., *Simulating Minds*. Oxford, Oxford University Press, 2006, pp. 238-9.
34. ENGELBERT M. and CARRUTHERS P., *Introspection*. *Wiley Interdisciplinary Reviews: Cognitive Science* 2010; 1: 246.
35. GOLDMAN A.I., *The psychology of folk psychology*. *Behavioral and Brain Sciences* 1993; 16: 15-28.
36. See CARRUTHERS P., *Simulation and self-knowledge*. In: CARRUTHERS P. and SMITH P.K. (eds.), *Theories of Theories of Mind*. Cambridge, Cambridge University Press, 1996, pp. 22-38; NICHOLS S. and STICH S., note 30.
37. GOLDMAN A., see note 33.
38. CRAIG A.D., *How do you feel? Interoception: The sense of the physiological condition of the body*. *Nature Reviews Neuroscience* 2002, 3: 655-66.

Troubles with self - consciousness

39. See ENGELBERT M. and CARRUTHERS P., see note 34, p. 247.
40. CARRUTHERS P., *The Opacity of Mind: The Cognitive Science of Self-Knowledge*. Oxford, Oxford University Press, 2011.
41. See, e.g., DEHEANE S. and CHANGEUX J.P., *Neural mechanisms for access to consciousness*. In: GAZZANIGA M. (ed.), *The Cognitive Neurosciences III*. Cambridge (MA), MIT Press, 2003, pp. 1145-58.
42. CARRUTHERS P., *How we know our own minds: The relationship between mindreading and metacognition*. Behavioral and Brain Sciences 2009; 32: 124.
43. BRASIL-NETO J.P., PASCUAL-LEONE A., VALLS-SOLE J., COHEN L.G., HALLETT M., *Focal transcranial magnetic stimulation and response bias in a forced-choice task*. Journal of Neurology, Neurosurgery, and Psychiatry 1992; 55: 964-6.
44. JERVIS G., see note 4, p. 157.
45. CARRUTHERS P., *Cartesian epistemology: Is the theory of the self-transparent mind innate?* Journal of Consciousness Studies 2008; 15(4): 28-53.
46. CARRUTHERS P., see note 42, p. 138. Here Carruthers is describing a hypothesis put forward in WILSON T., *Strangers to Ourselves*. Cambridge (MA), Harvard University Press, 2002.
47. JERVIS G., see note 7, p. 28.
48. JERVIS G., *Presenza e identità*. Milan, Garzanti, 1983, ch. 3.
49. JERVIS G., see note 48, p. 67.
50. BACON F., *Novum Organum*, 1620, Bk I, 41; Engl. trans. in *The Collected Works of Francis Bacon*. London: Longmans, 1857-74, vol. IV.
51. JERVIS G., see note 1, p. 348.
52. JERVIS G., see note 1, p. 302.
53. JERVIS G., see note 1, p. 301.
54. JERVIS G., see note 1, p. 301.
55. In the course of time Jervis has constantly recognized his debt to DE MARTINO: see, e.g., see note 48, p. 8; and see note 6, cap. 1.
56. DE MARTINO E., *Il mondo magico. Prolegomeni a una storia del magismo*. Turin, Einaudi, 1948, p. 158; Engl. trans. *Primitive Magic*. Dorset, Prism Press, 1988, p. 146.
57. JERVIS G., note 7, p. 92.
58. DE MARTINO E., *Storia e metastoria*. Lecce, Argo, 1995, p. 103; Engl. trans. in MASSENZIO M., *The Italian school of "history of religions"*. Religion 2005; 35: 214.
59. See SAUNDERS G.R., "Critical Ethnocentrism" and the Ethnology of Ernesto De Martino. American Anthropologist 1993; 95(4): 882.

60. DE MARTINO E., *Fenomenologia religiosa e storicismo assoluto*. Studi e materiali di storia delle religioni 1953-4; 24-25: 19; Engl. trans. in SAUNDERS G., *The Crisis of Presence in Italian Pentecostal Conversion*. American Ethnologist 1995; 22(2): 335-6.
61. JERVIS G., *Ricordo di Ernesto de Martino*. In: GALLINI C. and MASSENZIO M., *Ernesto de Martino nella cultura europea*. Napoli, Liguori, 1997, pp. 318-9.
62. KANT I., *Critique of Pure Reason*. KEMP SMITH N. (translated by London), Macmillan, 1929, B157. In B 158 Kant writes that “[t]he consciousness of self is [...] very far from being a knowledge of the self”.
63. JERVIS G., see note 48, p. 49.
64. JERVIS G., see note 7, pp. 131-2.
65. JERVIS G., *La conquista dell’identità*. Milan, Feltrinelli, 1997, p. 35.
66. JERVIS G., see note 7, p. 25.
67. JERVIS G., see note 65, p. 33.
68. “*In other words, being endowed with a self-consciousness [...] we all are prisoners of the uncertainties that concern it. And these are first of all uncertainties of image; secondly, and immediately after that, uncertainties of acceptability of ourselves ‘as we are’; and then – more distressing – uncertainties of consistency, that is, of inner solidity*” (JERVIS G., note 65, p. 33).
69. JERVIS G., see note 61, p. 319.
70. JERVIS G., see note 7, p. 92.

Correspondence should be addressed to:

Massimo Marraffa, Dipartimento di Filosofia, Università Roma Tre,
via Ostiense 234, 00144 Roma. E-mail: <marraffa@uniroma3.it>