1 Foreword

I wish to put in writing some observations on an intellectual matter I deem of special interest. One could perhaps dare to reduce it to the study of brains. In this sense, the matter is circumscribed to the inquiry of what may be termed a "representative instinct" in mankind. Such "instinct"—I use the term loosely for now—is plausibly the byproduct of the emergence of symbolic faculties in instinctive beings. This may not be too precise, but it is also not blatantly wrong, and thus I accept it as a decent preliminary formulation. I am speaking of what a peculiar branch of psychoanalysis termed *archetypes*.

The object of this inquiry, I must say, was not originally conceived to be of biological nature, and I do wish to attend to the earlier formulations of the concept of *archetype*. Firstly, because in the genealogy of any scientific idea there are large branches of—in general quite illuminating—pre-scientific antecedents. Secondly, because the hypothesis that concerns us goes far beyond the scope of scientific inquiry, raising questions about human nature itself.

As a last preliminary comment, I should wish to say the following: although the concept, as we postulate it, has a tighter link with scientific understanding than its original formulations, and although a considerable amount of empirical work has been set forth by affective neuroscience on its favor, the existence of archetypes is still merely a hypothesis. Only the scientific study of brain functioning can ultimately provide satisfactory evidence in favour or against it. It is probably the case that mind-problems are of the kind where propositions are to be measured only in that way—by the weight of existing evidence, and not by conclusive proof. A love for sharp edges and rigor compels me to limit the scope of my speculation to what I deem to be in line with this evidence and our current scientific understanding of the brain. It is utterly short-sighted to claim that reason can only delve into matters where certain knowledge exists (or could exist). The advancement of science itself must always begin with educated conjectures, and a philosophical attitude towards uncertain matters can *at least*, when executed correctly, reveal the right and the wrong questions to be raised.

2 Archetypes

The term archetype dates back at least to Dionysius the Pseudo-Areopagite, who suggested an echo (*apekhémata*) of the divine essence exists on every sensible object, by virtue of which they may elevate to the immaterial *arkhitypía*. Here, the word expresses the perfect platonic ideal of which each sensible object is an imperfect realization. This is not the sense of the term that interests us—although the notion of *echo* (or *imprint, trace*, etc.) will be relevant. In fact, our issue is partly the different meanings the term ascribes to across traditions and contexts. Even the work of Jung, that elevated the term to an unprecedented intellectual dimension, lacks an unequivocal definition, sometimes confronting us with the suspicion that the very author bestows it with different meanings depending on the period of his intellectual life or the object of his exposition.

It is my opinion that this linguistic issue is not as daunting as it may appear at first. A simple reason is that many of the meanings commonly attributed to the term can be readily disregarded as nonsensical. Secondly, some ad-

vances in the field of neuroscience—particularly in the line of research set forth by Panksepp—have contributed a great amount of empirical material to the question. This rich set of facts lays out more plausible—and potentially falsifiable—notions of what may be meant by "archetype". In other words, though the riddle is far from answered, we have at our disposal a whole domain of reality that the ancient philosophers—or early psychoanalysts—lacked. Thus, the contending formulations, to our surprise, are in the end not very numerous.

It then becomes the question how to penetrate into the essence of the formulations at hand. For this purpose, aside from the consideration of empirical evidence, I intend to follow two elementary principles. Firstly, to assess the amount of presuppositions implied in each of them. Secondly, to think of their meaning as determined by the sensible effects the concept is represented to produce. In the latter lies whatever is to be meant by the concept; in the first, its intellectual economy.

3 Jung

Jung had an explicitly dualistic outlook on the psyche. He endorsed the philosophical stance according to which there is no evidence in favor of the hypothesis that links psychological phenomena to physical and chemical processes, that there is no reason to regard the mind as an epiphenomenon of matter, and that it should be treated as a *sui generis* factor—at least until the artificial creation of a mind can be established as an achievable endeavor. This is explicitly held in the work *Archetypes of the collective unconscious*, written around 1932. The state of evidence at the time may perhaps make this claim understandable. However, at least to my knowledge, he never explicitly relinquished it.

Jung frequently associated the notion of archetype to that of primordial image. This association is particularly present whenever he was interested in drawing the parallelism between mythological motifs and the archetypal phenomena he allegedly witnessed in his clinical work. He also ties the concept of archetype to the notion of pattern of functioning:

> Like every animal, he [the man] possesses a preformed psyche which breeds true to his species and which (...) reveals distinct features traceable to family antecedents. (...) We are unable to form any idea of what those dispositions or aptitudes are which make instinctive actions in animals possible. And it is just as impossible for us to know the nature of the preconscious psychic disposition that enables a child to react in a human manner. We can only suppose that his behavior results from patterns of functioning, which I have described as images. The term "image" is intended to express not only the form of the activity taking place, but the typical situation in which the activity is released" — 1959, pag. 78.

This evolutionary speculation is not trivial. It is in line with more recent expositions, such as those elaborated in the work of Campbell. However, it is a rather strange twist of the logical chain to propose an evolutionary basis for an immaterial phenomenon. The process of natural selection affects the course of biological species made up of material elements, all the way down to a rather peculiar acid, whose material nature—I should hope—is uncontested. Putting that

aside, to claim we do not know the mechanisms that make instinctive actions possible is false under the present state of science. The neuroscientific findings seem to support, to a certain extent, the Jungian hypothesis of "archetypal" behavioral patterns. We will come to discuss these findings later on.

To Jung, primordial images are contentless, structural patterns—just like an instinct, taken by itself, is also strictly formal. This is an important point to make wherever we find those delirious babblings mumbled by the sadly numerous followers of the new age philosophy, who distort Jung's theories to fantasize about a world of actual images, where this or that archetype "appears as" this or that other figure and contains some form of sacred message. Granted, at times the author made it easy for such misinterpretations to be drawn out of his work. But it is also true that Jung asserted above all *a.* the affective tone of these "primordial images", and *b.* that they should be understood not by virtue of an essential content but by their teleology—*id est*, the specific behavioral disposition induced by them.

> But we only arrive at the meaning of a physical organ when we begin to ask teleological questions. Hence the query arises: What is the biological purpose of the archetype? —1959, pag. 161.

It continues to be unclear in what way the biological conception of the archetype and the immaterial notion of the psyche may theoretically harmonize. But we shall leave this question aside for the moment to discuss a bit more deeply about these primordial images, or patterns of behavior.

With regards to their teleology, one mustn't too hastily convince himself that it exists. A fair number of biological traits are nothing but the byproduct of others. A general idea is that the phenomenology of archetypes—the specific behavioral dispositions that are understood to be archetypal—is the induction of an affective state by stimuli to which we were once selected to respond affectively, or stimuli resembling other to which we were selected to respond affectively, even when at the present state of history such response is unwarranted. To rephrase: In the same way the fact that sucrose was once scarce makes us feel a compulsive appetite for it even today, other stimuli are also imbued with, so to speak, archaic affect. In this regard, we find a clarifying and mundane example in *The Masks of God: Primitive Mythology*, by Campbell.

> Chicks with their eggshells still adhering to their tails dart for cover when a hawk flies overhead, but not when the bird is a gull or duck, heron or pigeon. Furthermore, if the wooden model of a hawk is drawn over their coop on a wire, they react as though it were alive—unless it be drawn backward, when there is no response.
>
> Here we have an extremely precise image—never seen before, yet recognized with reference not merely to its form but to its from in motion, and linked, furthermore, to an immediate, unplanned, unlearned, and even unintended system of appropriate action: flight, to cover. (...) Furthermore, even if all the hawks in the world were to vanish, their image would still sleep in the soul of the chick—never to be roused, however, unless by some accidente of art (...). With that the obsolete reaction of the flight to cover would recur; and, unless we knew about the earlier danger of hawks to chicks, we should find the sudden eruption difficult to explain.

> 'Whence', we may ask, 'this abrupt seizure by an image to which there is no counterpart in the chicken's world? (...)'.

It is not difficult to observe what is being suggested in this passage. Namely, that there were now non-present stimuli imbued with affect by virtue of evolutionary archaic systems; that the responses elicited by these stimuli were arguably propitious to survival; that these archaic responses can be evoked by "accidents of art", provoking seemingly unwarranted responses in people. The psychologist that risks to say this draws his attention, of course, to all those affective motifs deemed universal in human experience: patterns suspected to be ubiquitous even across different religious contexts and times, appearing with equal strength in uncommunicated cultures as in the spontaneous production of individual people.

The scope of this alleged universality is not altogether clear. Some have proposed general practices but not particular images to fall within it. Shamanism may be a good example. Others have drawn their attention to particular imaginations: for example, that of the universal flood. I do not wish to discuss this point, for I cannot see a way to circumscribe the speculation within a limiting frame. And where there is place for unbounded speculation there is no place for truth. So, having presented a general overview of the more or less pre-scientific conception of *archetypes*, I should wish to proceed with the equally interesting—but more epistemologically promising—attempts of neuroscience at tackling this question.

4 Archaic affect

Allow me to advance that I shall not give too much detail on the neurobiology of the findings here commented. Doing so would make this writing much longer and technical than I intend. I should rather wish to summarize what I understand to be the principal conclusions and postulates of affective neuroscience and their relationship to the concept of *archetype*. Before discussing these contributions, however, it is pertinent to say a few words concerning both their ontology and their epistemology. This is important because discussions concerning the mind are famously problematic, and it would be imprudent to simply delve into the neuroscience before addressing a few concerns.

With regards to ontology, it is a common mistake to believe that Descartes committed an intellectual sin when postulating his *res cogitans*. In fact, when doing so, he acted in accord with standard scientific practice. It is obvious that certain human faculties cannot be explained by mechanical laws. If matter was to be entirely described by the laws of mechanics, as was the stance at Descarte's time, then some other substance had to be postulated in order to explain these non-mechanistic faculties. When Newton and Galileo postulated *forces*, they were doing so in similar conditions. Rather than a positive scientific achievement, such postulate was a problematic, but inescapable necessity.

The standard philosophical stance today is a commitment to a materialist monism. I share this stance—but the conundrum which forced Descartes to reject monism remains. It seems that, rather than postulating a new substance, the appropriate position is a form of Spinozian *dual-aspect monism*, as Panksepp

described it. It is a fact that thought emerges from matter—it is happening now as I write—but it is plausible that properties emergent from matter are better understood in non-material terms. This is in no sense an ontological claim; I simply state that a complex system may be not be properly described by the same laws which describe its constitutive elements. So, it seems the missing piece which Descartes' *res cogitans* aimed to account for were, in fact, complex neural dynamics. *How* complex neural dynamics may produce thought, emotion, etc., pertains to scientific inquiry rather than to philosophical speculation. Affective neuroscience, in particular the work of Panksepp and Damasio, has provided an immense corpus of evidence in favor of the notion that conciousness is not only an emergence of complex neural organization, but particularly of those deep regions of the brain whose function seems to be the realization of emotion and affect.

With regards to epistemology, the principal question is: How can we study emotions scientifically? The answer to this question is given by three observations which are, to me, scientifically impeccable. *(1)* Animals exhibit outward indicators of emotional states. The clearest example are the separation distress calls. *(2)* We can inquire on the neurobiology regulating such expressions using standard scientific methods. *(3)* If artificial exposition to the neurobiological regulators of such emotions produces corresponding *feelings* in humans, the weight of the evidence favors the conclusion that animals also experience these feelings—at least when the brain systems involved are homologous. For example, in a paper entitled *Opioid blockade and social comfort in chicks* (1980), Panksepp showed that opioid blockade with naloxone reduced the imprinting effect in chicks. More generally, his findings at the time suggested that opioids mediate social interaction and bonding. More recent papers of his added a great deal of evidence to this claim. We know certain opioids, such as oxytocin, mediate human bonding and partially regulate the feelings of warmness and care present, for instance, between a newborn and a mother. Since the neurobiological circuitry involved is homologous in both species, it is not unreasonable to claim that similar feelings permeate the bond between a little chick and its mother.

As a last note, and to put this preamble aside, I should say that I follow Panksepp's terminology. By *emotion*, I mean any affective, cognitive, behavioral or physiological change in the organism. In this sense, the word is more tightly linked to its etymology than to its conversational meaning. By *affect*, I mean quite broadly any subjective and experiential feelings, whereas an *emotional affect* denotes the experiential component of an internal brain state.

Emotional affects are generally associated to external events. When compared to external inductors of affect, internal ones (e.g. the memory of a deceased loved one) are rather few, and perhaps limited to our own species. Even so, every affect is an internal function of the brain. Scientists have used the terms *valence, arousal* and *surgency* to describe different aspects of the affective experience, without ever daring to speak of emotions. The present state of science, so I believe, conclusively leads to the fact that such terms would make no sense if not denoting affect.

The essential thesis of affective neuroscience as a field, as far as I understand it, is the following: that affective experience "reflects a primitive form of

consciousness which was the evolutionary platform for the emergence of more complex forms of consciousness" (Panksepp, *Affective consciousness: Core emotional feelings in animals and humans*, 2005). Here, *consciousness* refers to "brain states that have an experiential feel to them, and it is envisioned as a multi-tiered process that needs to be viewed in evolutionary terms, with multiple layers of emergence." (*idem*).

The evidence in favour of this thesis is close to overwhelming and, to my eyes, at least in essence, little doubt remains about their truth. There are, of course, disputes concerning technicalities—the manner in which different ways of consciousness evolved, the neurobiology modulating certain aspects of consciousness—but the essential claim remains untouched. Not only is the experiential component of emotion common to, at least, all mammals, but it represents a primitive form of consciousness that is still present in the human brain. This presence is attested by the fact that the systems from which such form of consciousness emerges remain very well preserved and highly homologous across mammal species. It is this evolutionary insight what draws a link between the conclusions of affective neuroscience and pre-scientific speculations on psychological archetypes.

Panksepp used the term *equalia* (*e* from evolutionary, *qualia* from its traditional philosophical sense) to denote the varieties of positive or negative feelings not simply mediated by our perceptual interfaces, but by the orchestration of inherited emotional systems in the brain. The raw affects engendered by such subcortical systems are "ancestral memories (instincts) that promote survival" (the reader is surely reminded of Campbell's example of the chick and the eagle).

It is not hard to see that these "ancestral memories" are what Jung termed *primordial images*. Needless to say, a highly complex, and in many senses still impenetrable interplay exists between the archaic emotional systems of the mammalian brain and the cortical regions to which our own species owns its so-called superior faculties. Notwithstanding, these cortical regions are entirely inessential to primary forms of consciousness; such forms subsist even in extreme cases of virtually non-existing cortical development. This was proven not only in Panksepp's work, but quite famously by Damasio as well.

Indeed, it seems that consciousness is like a curl or (in Spanish) *rizo*; the superior aspects of it can be removed while preserving pre-conscious affect, but whenever its primitive components are lost, nothing of it subsists. The different "levels" of consciousness have been described in the neuroscientific literature as *primary, secondary* and *tertiary* consciousness; or as *core consciousness* and *extended consciousness*. Regardless of the terminology, lower levels of consciousness have been demonstrated to depend on the primitive, value-encoding neurocircuitry described before. In short, consciousness can exist without cognition—but it cannot be without affect.

The relationship between the neurocircuitry of affective experience and cognitively complex phenomena, such as spirituality or dreams, has not been explored so far. However, in so far as affect is, virtually by definition, value-encoding, it seems necessary that spiritual experience is some kind of cognitive sublimation of these archaic components of human experience. This would at least account for *a.* the striking structural similarities across the wide range of

so-called archetypal stories and beliefs—the hero story, animism, shamanism, etc.—*b.* the profound affect with which such stories and beliefs are imbued; and *c.* the absolute resilience of such narratives in a scientific society. The same can be said of dreams.

There is one peculiarly interesting paper, (The affective Core of the Self)[https://www.frontiersin.org/artic which surveys the insight provided by affective neuroscience into one of the specific archetypes proposed by Jung. I will let the reader judge the conclusions of this paper on his own; they do not differ from what I have described thus far but in scope and technical precision.

5 Summary

Let us return to Campbell's quote:

> Chicks with their eggshells still adhering to their tails dart for cover when a hawk flies overhead, but not when the bird is a gull or duck, heron or pigeon. Furthermore, if the wooden model of a hawk is drawn over their coop on a wire, they react as though it were alive—unless it be drawn backward, when there is no response.
>
> Here we have an extremely precise image—never seen before, yet recognized with reference not merely to its form but to its from in motion, and linked, furthermore, to an immediate, unplanned, unlearned, and even unintended system of appropriate action: flight, to cover. (...) Furthermore, even if all the hawks in the world were to vanish, their image would still sleep in the soul of the chick—never to be roused, however, unless by some accidente of art (...). With that the obsolete reaction of the flight to cover would recur; and, unless we knew about the earlier danger of hawks to chicks, we should find the sudden eruption difficult to explain. 'Whence', we may ask, 'this abrupt seizure by an image to which there is no counterpart in the chicken's world? (...)'.

This is indeed the mysterious fact unraveled by affective neuroscience. For all practical purposes, it is correct to say an *ancestral memory* exists in the chick; one deeply embedded in ancient subcortical systems that mediate affective experience. Campbell presented his example to represent in practical terms the meaning of *archetype* in Jungian psychology. I think it is correct to say the existence of archetypes, *precisely* as understood by Jung, counts today with a substantial amount of evidence, and fits entirely with the advancements of neuroscience. Somewhere in his complete works—I remember not where—Jung states that we are, so to speak, *thousands of years old*. This is sensibly true. To speak of the value-encoding circuitry which lays well beyond cognitive control, in the deepest regions of our brain, as a *personality*, as Jung did, is merely a verbal matter.

Pseudo-Areopagite, I repeat, suggested a celestial echo (*apekhémata*) exists in every sensible thing; and that by virtue of such imprint they may reach the immaterial *arkhitypía*. The *ancient memories* revealed by neuroscience, quite curiously, have something of both of these Greek concepts. They are *apekhémata* insofar as they are an ancestral trace, an evolutionary imprint that, so to speak, echoes through time by means of millennial evolution. They are *arkhitypía* in that they exist deep beyond the cognitive regions of our brain, evoking typical

patterns of behavior across a range of biological species, while being perhaps impossible to represent by means of pure cognition or imagination.