10. What Mary was not Conscious (of). Scientific and Philosophical Perspectives on Consciousness in the light of the Knowledge Argument

1. Consciousness in Contemporary Sciences

Consciousness has not always been a matter of scientific interest. The sentence itself indirectly shows the reason why, for a long time, the study of consciousness has been an exclusive privilege of philosophy. Indeed, it seems paradoxical to think about consciousness as a "matter" of scientific research, because of its empirical evanescence and qualitative – more than quantitative – features, which make apparently impossible to catch such a phenomenon with a scientific method. Actually, up until a little more than a century ago it seemed impossible to empirically see an electron or, better still, to discover the elementary particles of the universe; however, nowadays we know that it was only a contingent impossibility, progressively dissolved by the scientific evolution. In general, what for a long time has been a mere matter of theoretical and conceptual analysis of philosophy, has become, over time, an object of experimentation. Just consider the revolution from philosophical theories about the nature of the universe to the scientific determination of the law of physics.

Hence, it seems consistent to say that also consciousness might be a potential matter of scientific study, thanks to new methodologies, new experimental disciplines – e.g. neuroscience and cognitive sciences – and new experimental instruments. The huge increment of neuroscientific researches about consciousness is a proof of this belief[1]: in the last decades, neuroscientists, besides focusing on perceptual and cognitive abilities – and impairments – of the brain, they also began to deal with the concept of consciousness as such, trying to find it within the empirical and experimental domain of cognitive sciences. Consequently, they started to talk about 'visual consciousness', 'perceptual consciousness', 'level and content of consciousness', 'awareness' and 'attention' as strictly related to consciousness – even not identical to it.

However, consciousness is a very general and vague term and there is not a univocal meaning for it; it seems more like an umbrella concept with a great range of

[1] Medline trend for 'Consciousness' (year: number of citations): 1996: 647; 2006: 1280; 2016: 2255. URL = <http://dan.corlan.net/cgi-bin/medline-trend?Q=Consciousness> [Accessed September 2017]

significances (Van Gulick 2017). Indeed, consciousness could be interpreted as sentience – i.e. the ability of sensing and responding to the world; as wakefulness, that is the awake and alert state in opposition to sleep, coma or other altered states; as self-consciousness or the fact that 'I am aware that I am aware'; even as the 'what it is like to be' state – i.e. the subjective notion of being conscious that makes a mental state conscious as such; and last but not least, as transitive consciousness or the fact that we are always conscious "of something". Moreover, it is possible to use the concept of consciousness not only for humans, but also for animals and even for potential artificial intelligence in the future.

So, there are many ways to intend the concept of consciousness, and there are also many methodologies to approach it, e.g. phenomenology, introspection, qualitative differences and dissociation methods, neurophysiological and behavioural methods and others. Regarding this methodological richness, it seems that the definition of the concept of consciousness seems extremely related to the kind of method we use for: any methodological approach seems to refer about a phenomenon very different from other types of measurement. For example, the discordances between the direct and rich visual perception of the surrounding environment we have each day and the results from the most recent findings from cognitive sciences about our not-so-complete visual experience (Dennett 2006, 82-84; Levy 2006, 40-42; against this idea, cf. Haun et al. 2017). There is a lack of convergences among different measurements and analysis (even between methods from the same epistemic source – i.e. theoretical or empirical ones), and, therefore, it is not clear if the result of all these different methods refers to a single common phenomenon or to a variety of conscious phenomena. In this sense, it might not be so easy to develop a genuine, empirical method for a scientific study of consciousness[2].

2. The epistemological roots of Contemporary Sciences

I shall claim that these methodological and definitional problems are only symptoms of a rooted epistemological issue, related to the structure of science as such. In fact, if we turn back to the rise of the modern science as we know today, we can notice that at its very epistemological ground there is no room for consciousness. In the seventeenth century Galileo Galilei gives birth to the scientific method, consequently producing a reformulation of the relation between subject and object. According to him,

[2] According to Elisabeth Irvine, the only way to solve this problem of misidentification is to eliminate the concept of consciousness from the scientific vocabulary; see Irvine 2013, 86, 173.

a theory is scientific if founded on «sense experiences and necessary demonstrations» (Galilei 1615/1897). The sense experience is the empirical perception of the world, against any aprioristic dogmatism on natural facts. While, the necessary demonstration is the elaboration of observations within a mathematical framework, in order to avoid any uncontrolled and non-rigorous empirical collection. So, the 'Novissima scienza' becomes an interdependent integration between experimental data and the mathematical model; and this integration is unlimited, towards an increasingly perfectible approximation (Zanin 2011). In other words, with mathematics, the nature becomes measurable, reproducible and predictable.

In order to do so, Galilei overturns the classic Aristotelean opposition between subjective and objective properties, eliminating the former from the scientific domain. Focusing on the conscious perceptiveness, Aristotle distinguishes between the 'proper sensible', and the 'common sensible': the former is the set of objective properties really possessed by the perceived thing, which can be directly sensed – e.g. colour, sound, taste, etc.; the latter is that kind of properties that can be indirectly sensed by more than one sense together – e.g. size, motion, weight, unity, number, figure – and these, due to their indirect derivation, they are capable of subjective error (Aristotle 2001, B6, 418a 16-19). Galilei overturns this opposition and what he calls 'primary qualities' – Aristotle's common sensible – become real properties of the world, while 'secondary qualities' – Aristotle's proper sensible – become product of the perceiving subject. Hence, without the perceiving and subjective act of knowledge, secondary qualities do not truly exist and, in this way, Galileo limits the domain of scientific knowledge, excluding from the natural world secondary qualities (Galilei 1623/1995). Or better yet, according to the Galileian paradigm, the only way for the pursuit of knowledge to be 'scientific' – i.e. rigorous and objective – is to exclude all the subjective experiences from its research program. Consequently, this lead to a structural impossibility for the objective knowledge to approach consciousness, understood as the subjective matrix of the whole experience[3]. Thus, long before the IXX century and the defence of the autonomy of phenomenal events and the contemporary philosophy of mind as such, the mental and the conscious experience have been abolished from the epistemological domain of science, as we know nowadays.

[3] The impossibility to picture subjectivity as part of our world view, because the subjectivity in question is the picturing representation. See Searle 1992, 98.

The repercussions are visible in the recent debate about the relationship between neuroscience and philosophy, focusing, in particular, on some epistemological and ontological issues, such as the Explanatory Gap Problem, the Knowledge Argument and the Hard Problem of consciousness (Levine 1983, Jackson 1982, 1986; Chalmers 1995). In very general terms, all three problems claim that any physical explanation will never be able to account for the emergence of conscious experience. In other words, the target of the argument is the alleged possibility to completely explain consciousness within a pure physicalist perspective and by means of a reductive approach to science. The main difference among the three is the degree of ontological commitment that any single thesis sustains against reductive physicalism: for example, while the Explanatory Gap Problem tries to prove the epistemic incommensurability between the function explained and the experience felt, the Hard Problem of consciousness exposes the idea that conscious experience involves non-physical properties – quoting David Chalmers, an "extra ingredient". I am not going to expose and explain these philosophical problem, nevertheless my intention is to highlight their common point, that is, the epistemological structure of contemporary science, that directly descends from the Galileian model of science. To do this, I am going to focus attention on the knowledge argument, that I interpret as a useful middle ground, with interesting fluctuation between pure epistemic and strong ontological claims. The aim is not to establish the consistency of the argument, but to sustain that, epistemologically speaking, some part of contemporary sciences does not walk a step forward from the modern scientific paradigm, and therefore we must reconsider the concept of science when we consider consciousness.

3. The Knowledge Argument

Frank Jackson first formulated the thought experiment in 1982, revisiting it in 1986 and thereafter, in order to reply to a great debate created around this issue (Alter 2007; Nida-Rümelin 2015). Here I quote the original version of the argument (Jackson 1982, 130):

> «Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. She discovers, for

example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue'. (It can hardly be denied that it is in principle possible to obtain all this physical information from black and white television, otherwise the Open University would of necessity need to use colour television.) What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false».

Essentially, the argument can be summarized in some passages (Nida-Rümelin 2015):

1) «Mary has all the physical information concerning human color vision before her release.
2) But there is some information about human color vision that she does not have before her release.
   Therefore
3) Not all information is physical information».

The argument claims that the mere physical information cannot explain all the phenomena in the world, and not for any contingent reason. When we say that Mary has all physical information, it should be intended as all the possible physical knowledge about the physical world – i.e. the past, present and future knowledge that a science grounded on physical causal closure and naturalism could potentially give to us. Arguably, Jackson's argument is not the first against the idea of completeness of physical knowledge – e.g. before him, Broad, Feigl, Nagel (Nida-Rümelin 2015). However, in the last decades of the past century the debate about the possibility or the impossibility to account for phenomenal states within the approach of Natural Sciences was very intense and heated. For this reason, the Knowledge Argument is so important, even today, where new neuroscientific discovers about the neural mechanism of our consciousness and cognition are a daily occurrence[4].

[4] For example, cf. Reardon 2017.

The ways in which we can interpret the meaning of the argument are variegate. The original intent had a genuine ontological valence, but in its development, as I already said, the argument could change its purpose, depending on the ontological commitment we want to sustain[5]. Assuming, for argument's sake, that the claim is consistent and coherent, it has to be discussed what is the meaning of "non-physical". Hence, the argument is mainly understandable in two ways: the epistemological (weaker version) and the ontological (stronger version) (Horgan 1989). The former sustains that there is a kind of knowledge which is not concerning physical, more similar to the Explanatory Gap Problem, while the latter that there are some non-physical facts *tout court*, as in the Hard Problem of consciousness. There are also more detailed versions of the argument, which try to develop all the shades and intermingling of epistemology and ontology, such as the distinction between metaphysical and linguistic physicalist, or the Ability Hypothesis and the Acquaintance Hypothesis, and the non-reductive physicalist position of the New Knowledge/Old Fact View (Nida-Rümelin 2015). I am not going to analyse all these variants, rather my aim is to evaluate and discuss a negative answer to the argument, with the purpose of consequently claiming my own position, which lies beyond the determination or refutation of argument's correctness. So, I shall expose Patricia Smith Churchland's analysis of the Knowledge Argument, taken from her notorious book *Neurophilosophy: Toward a Unified Science of the Mind-Brain* (1986).

4. Churchland's critical analysis of the Knowledge Argument

In the chapter 8 of her book, Patricia Churchland aims to demonstrate the reducibility of mental states to neurobiological ones, by analysing and discussing some of the most important position that sustains, by contrast, the irreducibility of them: Substance and Property Dualism, the 'What it is like to be' Argument from Nagel (1974) and Jackson's Knowledge Argument. Churchland generally sustains the possibility of a forthcoming complete intertheoretic reduction of psychology to neuroscience. Hence, if Mary has «the utopian neuroscience at hand», but she does not know what it is like to see

[5] Searle 1992, 117: «I think the argument is decisive, though it is frequently misunderstood in ways that treat it as merely epistemic and not ontological. lt is sometimes treated as an epistemic argument to the effect that, for example, the sort of third-person, objective knowledge we might possibly ha ve of a bat' s neurophysiology would still not include the first-person, subjective experience of what it feels like to be a bat. But for our present purposes, the point of the argument is ontological and not epistemic. lt is a point about what real features exist in the world and not, except derivatively, about how we know about those features».

the redness of a tomato – i.e. the specific living experience of red – it means that even the utopian neuroscience leaves something out, therefore it will be impossible to reduce psychology. In other words, her aim is not to negate the existence of some conscious events – e.g. the perception of redness, of a waterfall or the unpleasant pain made by a needle puncture – but she argues for the identity between folk psychology's content – i.e. perceptions, beliefs, feelings, and so forth – and brain states (Churchland 1986, 334): just as light is electromagnetic radiation, so conscious experience is a brain state. In this way, neuroscience, as a more rigorous and empirically coherent discipline and kind of knowledge, should reduce psychology into its domain. According to this point of view, the Knowledge Argument is necessarily wrong, because it leads to a differentiation between the facts we know about colour vision through the utopian neuroscience and the perception of redness.

First, as Paul Churchland and David Lewis has already argued (Jackson 1986), Patricia Churchland claims that there is a difference between the meaning of 'knowledge' used in the two premises – i.e. (1) Mary knows everything there is to know about brain states; and (2) It is not the case that Mary knows everything there is to know about sensations and their properties. (1) involves propositional knowledge (manipulation of concept from all the possible physical knowledge), while in (2) there is more a kind of pre-linguistic apprehension, maybe involving «innate disposition to make certain discrimination». According to Churchland, the argument lies on this equivocation and therefore it does not really work. Indeed, there could be two or more ways of knowing the same thing in the world, without sustaining the failure of physicalism. She uses pregnancy as an example: the process known by a childless obstetrician – with a kind of propositional knowledge (1) – is the same process known by a pregnant but untutored woman – mostly with a kind of pre-linguistic knowledge, made of the interoceptive sensitivity that permits to feel the baby growing into her womb. At the same way, the redness known by type (1) – e.g. by neurophysiology of perceiving colours and (2) of knowledge – e.g. the mere act of seeing a tomato – is the same object or property. The fact that Mary does not perceive (2) red when she is studying all the physical facts related to our colour perception (1) is not more problematic than the fact that the obstetrician does not become pregnant by knowing all the physical facts about pregnancy.

The second point against the argument is focused on the fact that Mary, after her release, knows something new in addition to all the physical facts already known by means of the utopian neuroscience (Churchland 1986, 332-333). According to

Churchland, this idea seems obvious and intuitive only *prima facie*, because it is impossible to know something new, if the first premise says that Mary knows everything there is to know about the brain, everything from the utopian – not actually existing – neuroscience. As far as we know, our current neuroscience might be for the utopian neuroscience what was Aristotelean physics for the current theoretical physics. A complete neuroscience should be able to reach all the facts related to the brain, also via introspection. For example, once released Mary from her white and black laboratory and showed her a red tomato, she could see, through an introspective way given by utopian neuroscience, that she has, so to speak, a precise x state in her y pattern, and she knows that having that state is identical to perceiving red. If the premise of the argument assume that she knows everything about the brain and its properties, so this example could be possible in principle, and claims that this is impossible because brain states and introspective mental states are not identical leads the argument to a *petitio principii*.

5. The "Pain Argument"

I agree with the Churchlands and Lewis when they argue that there is a different use of the concept of knowledge in the two premises of the argument, but in any case, an objection to Patricia Churchland's first claim can be made as follows. I see no problem in saying that there are (at least) two different ways of understanding and using the concept of knowledge, and I sustain that, at most, we have to consider the knowledge argument only in its epistemological value. Churchland's argument can be interpreted within the New Knowledge/Old Fact View, according to which Mary can make new items of knowledge (the redness of the tomato), but they are just physical facts already known before the release, under another conceptualization – i.e. phenomenal conceptualization. I think that Churchland underestimate the importance that a vast spectrum of epistemic possibilities can be for the understanding of the whole conscious experience. Regarding the example of pregnancy, it is true that, the process known by the childless obstetrician and the untutored woman is the same, physically speaking and independently from any knowing and perceiving subject. However, references could be very different with a double epistemic access of this kind, and they might imply a sort of non-directly physical properties – e.g. pain – that are not mere particular cases of demonstrative beliefs (contrary to Perry 2001), but a constitutive part of the determination of the event as such.

Let us consider the following situation: on one hand, there is the most distinguished obstetrician in the world – male, affected by CIPA[6] – and on the other side an untutored pregnant woman, as in the previous example. The obstetrician, without doubt, knows everything there is to say – physically, chemically, cerebrally speaking, etc – about the process of pregnancy, and, obviously, even all the complications arising from the childbirth. Right in that moment, while the woman is fully in her grief, the obstetrician says: "I can imagine the pain, now please try to relax"; and suddenly the woman snaps at him: "You have no idea what I'm feeling!". The woman's sentence is not to consider simply as a figure of speech, because the obstetrician can know all about the mechanisms producing the pain of childbirth, but without the direct reference of the pain felt, it does not have the complete comprehension of the event's meaning[7]. If we were talking about the process of chlorophyll photosynthesis, then the first-person experience on the mechanism for the absorption of light by chlorophyll might be a non-necessary demonstrative belief, and it cannot add nothing to the comprehension of the process as such. The difference with the previous example is that, regarding the case of pain during childbirth, any subjective and phenomenal feature is constitutive of the event as such, therefore, a physician could know all the physical process of pregnancy and childbirth, but he does not know the event in all its essential aspects.

I have stressed the thought experiment figuring the male obstetrician with CIPA, but I think that even a childless female obstetrician cannot completely understand the stress of labour, but can actually imagine it, at the same way if Mary saw many colours – but not red – before the release, we might say that she may be able to approximately imagine the redness of tomato. However, the situation of 'black & white' Mary, regarding colours, is more similar to the male obstetrician with CIPA, regarding the event of childbirth. Someone may argue that the mere description of all the physical facts related to the mechanism of vision or to the pregnancy stages are sufficient for a scientific explanation of their respective events. However, I may counter with the fact that, for some kind of events in the world, the phenomenal level of description – i.e. the subjective and qualitative level – is not complementary or juxtaposed to the scientific – objective,

[6] Congenital insensitivity to pain with anhidrosis, a rare disorder of the nervous system. People with CIPA are unable to respond to changes in the physiological conditions of all tissues of the body. Patients exhibit insensitivity to both superficial and deep painful stimuli, including visceral perception, but touch, vibration and position senses are normal (Indo 2010).

[7] I prefer to use the word 'event' instead of 'fact' because it refer more to a thing understandable under many perspectives, not only the chemical and physical ones. See Northoff 2004, 112.

naturalistic level – rather it is constitutive in the interest of a complete description and comprehension. Taking once again the example of pain, it seems that this kind of event especially needs a phenomenological determination, even for being considered ontologically existent.

According to Sartre, pain (no less than pleasure, fatigue, grief, joy, etc.) is lived as an immediate, 'non-thetic' kind of conscious experience, and, in order to explain this, he often indicates the consciousness of pain by putting the word 'of' into parenthesis – i.e. the so called non-positional consciousness (Sartre 2013, liii)[8]. His aim is to highlight the concreteness of the painful experience, against any interpretation of it just as a mere representation or some epiphenomenal qualia: there is not a consciousness first and then the receipt of pain, nor there is not first the pain as such and then the quality of 'conscious' for the painful state itself[9]. Certainly, pain is also an intentional and thetic object of our consciousness – i.e. positional consciousness[10] – therefore we can reflect on it as a psychic object transcendent to consciousness itself – e.g. concerning the relationship with our body or scientifically analysing causes and symptoms. However, to paraphrase Sartre himself, 'consciousness (of) pain' is constitutive of the pain as the very mode of its own existence, as the material of which it is made, and not as a form which is imposed by a blow upon a material[11].

6. The Appearance-Reality Fallacy

At this point, returning to Mary's question, we need to understand if this argument can be used also for colour knowledge. At first blush, the knowledge about colours seems more similar to the description of natural phenomena (light, heat, chlorophyll photosynthesis, hurricanes, etc.) made by science. If colour can be explained with the mere description of its physical component, without the introduction of a direct subjective experience, it follows that colour does not involve any other fact outside of physical facts

[8] Rowlands 2001, 144: «Every positional conscious act is also essentially characterised by a non-positional consciousness (of) itself, where the parenthesised 'of' indicates that nonpositional consciousness is intrinsic (*immanent*) to the state itself and does not possess the sort of extrinsic, relational, character characteristic of reflective consciousness (e.g. introspection). And Sartre is very clear that the concept of knowledge, being extrinsic and relational in character, is entirely inappropriate for capturing or characterising the concept of pre-reflective consciousness».

[9] Sartre 2013, 125: «[N]othing can separate the consciousness (of) belief from belief, since belief is nothing other than the consciousness (of) belief».

[10] Positional consciousness «transcends itself in order to reach an object» to which I direct my attention (Sartre 2013, xxvii).

[11] Sartre uses this words for the example of consciousness (of) pleasure, but phenomenologically speaking it works also with the consciousness (of) pain (Sartre 2013, 14; Reisman 2007, 28).

understandable by Mary, before her release. In *The Rediscovery of the Mind*, Searle claims that there is an ontological – not just epistemological – difference between, for example, the conscious state of pain and the perceptible property of redness (Searle 1992, 114-118). The red colour can be defined in an ostensive way with a series of examples, but if we causally reduce the phenomenon to his physical composition – e.g. referring to photon emissions of 600 nanometres – then we can «carve off and eliminate the subjective experience of color from the "real" color. Real color has undergone a property ontological reduction to light reflectances». On the contrary, we have already seen that this is kind of reduction is not possible with some conscious state – e.g. pain, pleasure, etc. – and even, Searle says, with consciousness as such (*Ibid*, 116):

> «But now we come to an apparently shocking asymmetry. When we come to consciousness, we cannot perform the ontological reduction. Consciousness is a causally emergent property of the behavior of neurons, and so consciousness is causally reducible to the brain processes. But – and this is what seems so shocking – a perfect science of the brain would still not lead to an ontological reduction of consciousness in the way that our present science can reduce heat, solidity, color, or sound».

Such kind of thesis can be found also one of the most important publication of Gerald Edelman and Giulio Tononi, *A Universe of Consciousness* (2001, 12-13):

> «Scientific explanations can provide the conditions that are necessary and sufficient for a phenomenon to take place, can explain the phenomenon's properties, and can even explain why the phenomenon takes place only under those conditions. But no scientific description or explanation can substitute for the real thing. We all accept this fact when we consider, say, the scientific description of a hurricane: what kind of physical process it is, why it has the properties it has, and under what conditions it may form. But nobody expects that a scientific description of a hurricane will be or cause a hurricane.
> Why, then, should we not apply exactly the same standards to consciousness? We should provide an adequate description of what kind of physical process it is, why it has the properties it has, and under what conditions it may occur. As we shall see, there is nothing about consciousness that precludes an adequate scientific

description of the particular kind of neural process it corresponds to. What, then, is special about consciousness? What is special about consciousness is its relationship to the scientific observer. Unlike any other object of scientific description, the neural process we are attempting to characterize when we study the neural basis of consciousness actually refers to ourselves—it is ourselves—conscious observers. We cannot therefore tacitly remove ourselves as conscious observers as we do when we investigate other scientific domains»[12].

The theoretical strategy is approximately the same both in Searle and Edelman: at the heart of the argument there is a malleable distinction between two types of ontological domains. On one hand, there is the domain of the 'surface features' and the 'appearance' (Searle 1992, 119) or the 'observer-related thing'[13] – on the other hand, the domain of the underlying reality, the 'real thing' as such. The first, the appearance's domain, may include both secondary and primary qualities – e.g. the subjective perception of red and the property of solidity[14] – and it may be progressively excluded from the scientific explanation, and substituted by the real domain, in which their underlying ontological causes are determined. The problem with consciousness and specific conscious states is that regarding that kind of explanation, our interest is necessarily focused on the subjective experience as such. In other words, the distinction between appearance and reality does not work with consciousness as the object of the investigation: «Where appearance is concerned we cannot make the appearance-reality distinction because the appearance is the reality» (Searle 1992, 122) and in such cases «no scientific description or explanation can substitute for the real thing» (Edelman & Tononi 2001, 12).

Now, I shall argue that, although it could suggest a right way to understand the issue of consciousness, this kind of epistemological strategy and ontological distinction,

[12] See also Edelman 2003, 5521: «To expect that a theoretical explanation of consciousness can itself provide an observer with the experience of ''the redness of red'' is to ignore just those phenotypic properties and life history that enable an individual animal to know what it is like to be such an animal. A scientific theory cannot presume to replicate the experience that it describes or explains; a theory to account for a hurricane is not a hurricane. A third-person description by a theorist of the qualia associated with wine tasting can, for example, take detailed account of the reported personal experiences of that theorist and his human subjects. It cannot, however, directly convey or induce qualia by description; to experience the discriminations of an individual, it is necessary to be that individual».

[13] Edelman & Tononi 2001, 13: «Unlike any other object of scientific description, the neural process we are attempting to characterize when we study the neural basis of consciousness actually refers to ourselves—it is ourselves—conscious observers».

[14] As well as secondary qualities, «[t]he same pattern works for the primary qualities: Solidity is defined in terms of the vibratory movements of molecules in lattice structures, and objective, observer-independent features, such as impenetrability by other objects, are now seen as surface effects of the underlying reality. Such redefinitions are achieved by way of carving off all of the surface features of the phenomenon, whether subjective or objective, and treating them as effects of the real thing» (Searle 1992, 119-120).

thus developed, actually leads to a misinterpretation of all the questions we are talking about in this paper. The reductive scientific view as well as the main conceptual paradigm of contemporary philosophy of mind, even the Knowledge Argument itself, are all rooted on this problematic ground. The problem raised by Searle and Edelman that is widespread in most recent discussions about consciousness, subjectivity and the relationship with the world, is that there is no such a truthful distinction between appearance and real thing – i.e. between subjective, surface properties and objective, scientifically-defined thing (Hales 2014, 86). As both Searle and Edelman sustain, scientific research should progressively pursue the determination of the real thing by excluding from the explanation all subjective contents – e.g. the real red as photon emissions as well as the real description of fluidity as a particular disposition of molecules, and so forth. On the basis of this kind of distinction, it follows that consciousness is a very strange phenomenon, because it seems impossible to remove completely the subjective point of view – epistemological version of the problem – or the subjective content of the experience – ontological version of the problem; I shall demonstrate that the problem is not this impossible removal about consciousness just described, but the paradigm on which the issue is developed. This is the basic paradigm on which the Knowledge Argument lies, because the problem is conceptually the same, that is, is a pure and complete knowledge of physical facts sufficient for the description of the nature of colour? In other word, within the distinction reality-appearance is it possible to achieve a physicalist knowledge that could fulfil every request of complete explanation of a phenomenon, without entailing any apparent and non-physical facts.

We have discussed the argument from the original version (the problem of redness) and from another peculiar example (the nature of pain), but I may argue that we could go on with examples of any kind. Colour and pain were useful in order to highlight the constitutive presence of the first-person perspective, not only for a complete description of both phenomena, but also for their existence. However, I could say the same about other phenomena that apparently do not require for a complete knowledge any non-physical, subjective or observer-related content in general. Let us take the example illustrated by Edelman and Tononi, the hurricane. They say – correctly – that a scientific explanation can provide the necessary and sufficient conditions for the

hurricane to 'emerge'[15], but it cannot substitute the real thing nor cause it[16]. Ontologically speaking, if we put in a supercomputer all the information, all the contextual contingences and all the variables carved from the analysis of a hurricane, no one would expect to witness the arise of a real hurricane; epistemologically speaking, not a single scientist would require from his colleagues the description of what it is like to be inside a hurricane for giving an exhaustive scientific description of the event – lucky for them. So, in short, we do not need any subjective features for define the nature of the real thing, rather we need to remove ourselves as conscious observer for a scientific explanation as such (Edelman & Tononi 2001, 13). And this is when the problems arise. The fact is that it is not true that we remove the observer-relation from the 'normal' objects of scientific research. Not only with consciousness, but also with every phenomenon studied by science we do not remove first-person perspective. We just pretend to analyse any process, event or fact with an objective point of view, understood as an 'observer-independent' perspective (Hales 2014, 30). The so-called third-person perspective is often considered as the most 'real' view, through which we can consequently approach the real thing, beyond all the appearances. Despite this assumption, we never get rid of the first-person perspective, on the contrary it is in every sentence and it is the ground zero of every sentence about the world. Instead, the third-person perspective is not the detector of an alleged underlying reality made of physical facts, but rather it is a kind of abstract agreement that it may become independent of the particular and contingent observer – e.g. the single researcher – but not of the observation as such (*ibid*, 87). This is not a critique to the scientific method as such, it is a claim against the idea of a division between a physical-objective-real domain and a non-physical-subjective-apparent domain.

7. The Phenomenological Priority of Consciousness

In this sense, consciousness is not a contingent point of view – e.g. the particular demonstrative case as in the Knowledge Argument – nor a non-physical object of the world with some undetectable features, nor a peculiar way of knowing the world. All these accounts entail some oppositions (interior-exterior; real-apparent; knowing-known, subject-object and so forth), and this kind of epistemological and ontological dualisms

[15] The concept of emergence is not used innocently, because it is one of the epistemological cornerstones on which most of contemporary scientific theory lies. See Searle 1992, 111-116.

[16] See Churchland 1986, 332: «Just as the obstetrician does not become pregnant by knowing all about pregnancy, so Mary does not have the sensation of redness by knowing all about the neurophysiology of perceiving and experiencing red».

are the product of the revolution of thought happened in the modern age, starting from Galilei and Descartes – respectively with the scientific method and the mechanistic view of the world. Even the opposition between first- and third-person perspective depends on the roots of modern philosophy. As Sartre says, trying to find behind the consciousness is the result of the illusion of 'worlds-behind-the-scene'[17], and this is what Searle, Edelman, Tononi, Jackson and so many other figures of contemporary neuroscience and philosophy of mind do.

Instead, if we acknowledge the impossibility to remove consciousness and its first-person perspective from experiencing the world as such, it follows that this truly ontological perspective is not to be understood as the beginnings of a solipsistic position, but as the series of phenomena, that is, what reveals itself as it exactly appears, with no hidden reality behind. On the contrary, by taking this phenomenological approach, the third-person perspective is a virtual view that could describe the world through abstractions of the 'lived world'[18]. I am not saying nothing anti-scientific, I am rather highlighting again the foundations of the scientific method, defined by Galileo Galilei, on which the scientific world still lies, but that maybe has misinterpreted. With the scientific method, Galileo did not divide the world in the real one and the apparent one, he only tried to analyse some portions of the world[19], by means of the idealization of the nature – i.e. the approximation of the real event into a model-event with the lack of subjective properties and non-measurable contingencies. Anything to do with the idea of a complete knowledge, made of mere physical facts and with the exclusion of useless subjective features. For this reason, also the Knowledge Argument falls into the appearance-reality fallacy and therefore it is so conceptually slippery and problematic.

[17] See Sartre 2013, xlv: « […] an electric current does not have a secret reverse side; it is nothing but the totality of the physical-chemical actions which manifest it (electrolysis, the incandescence of a carbon filament, the displacement of the needle of a galvanometer, etc.). No one of these actions alone is sufficient to reveal it. But no action indicates anything which is behind itself; it indicates only itself and the total series».

[18] See Husserl 1970, 48-49: « […] we must note something of the highest importance that occurred even as early as Galileo: the surreptitious substitution of the mathematically substructed world of idealities for the only real world, the one that is actually given through perception, that is ever experienced and experienceable – our everyday life-world. This substitution was promptly passed on to his successors, the physicists of all the succeeding centuries».

[19] What is susceptible to mathematical correction, in order to reach a perfectible mind-world correspondence (Zanin 2011).

8. Actualist and Objectualist conceptions of Consciousness

Now, we have to consider how to fit this phenomenological approach to the Knowledge Argument, and consequently to the actual debate about conscious experience, between neuroscience and philosophy. In order to do so, I shall refer to the distinction between 'actualist' and 'objectivist' conceptions of consciousness, made by Mark Rowlands in *The Nature of Consciousness* (2001). According to him, one of the peculiar features of consciousness is its hybrid structure, as it can be both object and act of experience (Rowlands 2001, 122). Rowlands bases his claim on Sartre's distinction (see above) between positional (thetic) and non-positional (non-thetic) consciousness. In particular, non-positional consciousness, in virtue of its non-reflective, non-intentional and non-dyadic features, is very close to be a transcendental condition of the act of consciousness, rather than a form of knowledge[20] or the awareness of an object (Rowland 2001, 131). The problem is that most recent discussion are dealing with the conception of consciousness as object, while they are progressively eradicating the importance of the act of consciousness as such. I am not just talking about the physicalist paradigm within neurosciences are moving nowadays, looking for the necessary and sufficient condition for the emergence of conscious states. Rowlands highlights how even the what it is like to have a conscious experience – one of the most important argument against scientific reductionism – is grounding on the objectualist conception (*ibid*, 136-140). Hence, I am not slipping into the dualist division between the knowing subject and the known object, and, as will be explained below, this is not a revised version of the 'what it is like to be' argument nor a new kind of definition of subjective qualitative states – i.e. qualia.

The paradigmatic position in which I am going to transfer the entire debate about consciousness arises beyond all the distinctions mentioned above, including them within a higher level of conceptualization in which they are abstract determination of a phenomenological perspective. Thus, the aim is not to refute the distinction between subjective and objective features in the world, nor the existence of some kind of differences between the propositional knowledge of physical, objective facts about the world and the qualitative, a-conceptual experience of some facts in the world. In other

[20] See Sartre 2013, xxx: «The necessity of syntax has compelled us hitherto to speak of the 'non-positional consciousness of self'. But we can no longer use this expression in which the 'of self' still evokes the idea of knowledge. (Henceforth we shall put the 'of' inside parentheses to show that it merely satisfies a grammatical requirement.)»

words, I am not claiming that there is no epistemological or ontological difference between hurricane and consciousness; actually, these types of distinction belong to the objectualist view, while I shall argue that for understanding better the problem of consciousness we should focus on its phenomenological condition of possibility on which depend both the ontological and the epistemological determinations of the phenomenal.

For example, as Rowlands points out, even the 'what it is like' – originally formulated by Nagel – is often developed within the objectualist interpretation. In his paper entitled *What it is like to be a bat* (1974), Nagel argues that a reductive explanation of consciousness is impossible, because of the subjective-related form of every conscious experience. He formulates a specific epistemological situation in order to corroborate the argument: even if we knew all the objective information about the bat's ability of echolocation, we would not be able to capture the main feature that determine consciousness as such – i.e. the subjective character of experience that expresses the what it is like to be in this certain conscious state. This is clearly an argument against any physicalist and reductive theory of consciousness, however it is often included it in the objectualist interpretation, even if its purpose is to defend the incommensurability of subjectivity. Indeed, the what it is like of conscious experience is treated as state which we can be 'conscious of', rather than the act of «directing of consciousness towards a quite distinct object» (Rowlands 2001, 136). At the same way, the Knowledge Argument, understood in the light of these interpretations, should be included in the objectualist one, and therefore it misses the essential point of consciousness. The argument raises the issue as to whether the concrete subjective experience of something already known within a complete physicalist way can be conceived, in turn, as an object of (new) knowledge. In this scenario, the 'what it is like' to see the red colour is conceived as an object of consciousness – or better, an object of knowledge – thus, the attention is directed towards the object's epistemic relation to consciousness. But, as Rowlands points out, certain aspects of conscious states cannot be included under the epistemic relation with what Sartre calls positional consciousness: this relationship is extrinsic and refers to the dyadic concept knower-known (*ibid*, 146). Instead, some features of consciousness are immediately related to the directing act of consciousness towards objects, and not to the object of the directing act. Or, better yet, some features are related to the directing act as a directing act as such and not as an object of consciousness (see the example of pain exposed above). Hence, the Knowledge Argument is grounded on the objectualist interpretation, since it focuses on what a conscious state can know, rather than how a

conscious state can know. The argument only suggests that there is something concerning consciousness that is incommensurable for a mere physicalist knowledge, but the problem is not physicalism as such, but the underlying objectualist perspective. Jackson created this argument for claiming that conscious experience involves non-physical things and properties. However, due to the objectualist interpretation, this argument cannot move one step ahead from the Galileian differentiation between primary and secondary qualities and the resulting definition of the epistemological limit of the scientific method.

In this sense, the distinction between actualist/non-positional versus objectualist/positional consciousness is not comparable to the subjective-objective dyad, risen during the period of modern philosophy and the birth of the modern science. Rather, we can include in the same objectualist conception both naturalistic and anti-physicalist theories of consciousness – i.e. the key *explanandum* of consciousness (Chalmers 1995) – while developing a different set of methods and interpretations, mainly related to the phenomenological approach, the aim of which is to analyse the conditions of possibility of any act of experience as such – i.e. the key *explanans* of consciousness. Now, the last issue of this paper is to define what kind of approach could engage and investigate the actualist fold of consciousness, and for this task I may suggest neurophenomenology. Thus, in order to corroborate this perspective, I have to return for the last time to Churchland's critique to the Knowledge Argument.

9. What kind of (neuro?) Science of Consciousness

Until now, we have discussed the first part of Churchland's analysis about the Knowledge Argument. In the second part, as exposed above, she claims that if Mary knew all the physical facts about colour perception, it would mean that there is not any a priori reason why this utopian neuroscience should not be able to give a complete knowledge of colours, even the experience itself (Churchland 1986, 333):

> «Given that she is supposed to know absolutely everything there is to know about the nervous system, perhaps she could, by introspective use of her utopian neuroscience, tell that she has, say, a gamma state in her 0 patterns, which she knows from her utopian neuroscience is identical to having a red sensation. Thus, she might recognize redness on that basis».

It is not my intention to criticize Churchland's argument as such – although there would remain a few points to be clarified[21]. I now assume that Churchland is right when she says that the question whether we will be able to recognize a colour with a sort of utopian neuroscientific introspection is an empirical issue. Therefore, it is not possible to answer with an a priori argumentation, as well as an ancient astronomer could not have affirmed that in the future it would have been possible to investigate the 'invisible' black matter in the outer space[22]. Churchland claims that the utopian neuroscience may be quite different to the existing neuroscience, in order to have the possibility to know everything there is to know about the brain[23]. And «[e]verything is a lot, and it means, in all likelihood, that Mary has a radically different and deeper understanding of the brain than anything barely conceivable in our wildest flights of fancy» (Churchland 1986, 332). In other words, we can expect an epistemological and conceptual revolution of neuroscience, as it happened with the modern physics in the last century.

I want to seize this idea and suggest that Churchland is right, if she means that this hypothetical neuroscience of the future will go beyond the actual epistemological limits imposed by the Galileian scientific method (anyway, I do not think that she was meaning a neuroscience of this kind). In this sense, a utopian neuroscience of this kind should understand and incorporate in the same system both third-person and first-person perspectives[24], that is, avoiding the dualism between physical facts and subjective

[21] For example, the concept of 'identical' is problematic. The correlation – assuming for the sake of argument a necessary and sufficient correlation – between the gamma state and the red sensation does not entail the identity, certainly not an indiscernibility, because the neural pattern and the subjective sensation are characterized by different epistemic properties. Indeed, even in this imaginary situation, Mary the utopian neuroscientist needs to perform a mediate, conceptual passage of correlation and synthesis between the primitive sensation of red and the recognition of it as that precise gamma state. In other words, she does not 'see' the gamma state as red, rather she sees red and she introspectively recognize the presence of a gamma state, then she makes a cognitive correlation between the two events, but it is not possible to define this kind of relationship as a strong identity. Thus, even a hypothetical 'Gary the illiterate' may have a more direct knowledge of the visual perception as red sensation, than the necessary intellectual passage from the union between the visual perception and the neural pattern to the recognition of the red sensation.

[22] Actually, it would be possible to counter with the Kripkean argument against the alleged identity between, for example, the C-fiber stimulation in the brain and the sensation of pain (Kripke 1980, 148-155). The concept of pain is not a 'rigid designator' – i.e. it does not refer to the same object in all possible worlds. For this reason, the relationship between C-fiber and pain is not a necessary identity, unlike the identity between heat and molecular movement, that is true in all the possible worlds. Indeed, anyone could think at the C-fiber stimulation without necessarily evoke the sensation of pain and vice versa. The same reasoning can be made with the relation between the gamma state and the sensation of red. However, for the purpose of my thesis, I shall go on with Churchland's argument.

[23] Searle draws similar conclusions; Searle 1992, 124: «Furthermore, when I speak of the irreducibility of consciousness, I am speaking of its *irreducibility according to standard patterns of reduction.* No one can rule out a priori the possibility of a major intellectual revolution that would give us a new-and at present unimaginable conception of reduction, according to which consciousness would be reducible».

[24] About the necessity to find a middle ground for first- and third-person data, see Northoff & Heinzel 2006.

experience, with no reduction of any kind – basically, a science that could solve the Knowledge Argument. For this purpose, I shall suggest neurophenomenology, since its aim – in the words of Francisco Varela – is to find «meaningful bridges between two irreducible phenomenal domains […] instead of finding 'extra ingredients' to account for how consciousness emerges from matter and brain» (Varela 1996, 340)[25].

In this sense, neurophenomenology is relatively a young discipline that attempt to integrate the phenomenological analysis of experience with the empirical experimentation on the biological systems – i.e. combining first-person data and the dynamical analysis of neural processes to study subjects (Gallagher & Zahavi 2007). Thus, regarding the Knowledge Argument, from a neurophenomenological perspective the problem is not to assess if a complete physical knowledge suffices any information about colour vision, rather it is to understand how this knowledge about the brain mechanisms of vision mutually relates with the concrete, subjective act of seeing something red. In other words, it is an attempt to reunite primary and secondary qualities without any reduction, and considering the conscious experience primary as an action pragmatically, socially and culturally embedded, rather than a mere object of empirical studies. There is still a long way to go for neurophenomenology – there is no shortage of critics[26] – but I think that it would be a valid alternative to the objectualist interpretation of consciousness widespread among neuroscience and philosophy of mind.

Federico Zilio

(University of Padua, PhD Student in Philosophy)

[25] See also p. 347: «This research programme seeks *articulations by mutual constraints* between the field of phenomena revealed by experience and the correlative field of phenomena established by the cognitive sciences. I have called this point of view *neurophenomenology*».
[26] See, for example, Bayne 2004; Dennett 2006.

References

Alter, Torin. *The Knowledge Argument*. In Velmans, Max and Susan Schneider. *The Blackwell Companion to Consciousness*. Malden Oxford: Blackwell, 2007.

Aristotle. 2001. *L'Anima*. Milano: Bompiani testi a fronte, Giancarlo Movia (ed.).

Chalmers, David J. 1995. "Facing up to the problem of consciousness". In *Journal of Consciousness Studies* 2 (3):200-19.

Churchland, Patricia S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.

Dennett, Daniel C. 2006. *Sweet dreams: philosophical obstacles to a science of consciousness*. Cambridge, MA: The MIT Press.

Edelman, Gerald M. and Giulio Tononi. 2001. *A Universe of Consciousness. how Matter Becomes Imagination*. New York: Basic books.

Edelman, Gerald M. 2003. *Naturalizing consciousness: a theoretical framework*. In Proc Natl Acad Sci U S A. Apr 29;100(9):5520-4

Galilei, Galileo. 1897. *Galileo a Madama Cristina Di Lorena, 1615*. Padova: Tip. Salmin.

Galilei, Galileo. 1995. *Il Saggiatore, 1623*. Lecce: Conte, Eugenio Garin (ed.).

Hales, Colin G. 2014. *The Revolutions of Scientific Structure*. Vol. 3. New Jersey ; London: World Scientific.

Haun, Andrew M., Tononi Giulio, Koch Christof and Tsuchiya, Naotsugu. 2017. "Are we underestimating the richness of visual experience?". in *Neuroscience of Consciousness*, Volume 3, Issue 1, 1 January 2017, 1-4.

Horgan, Terence E. 1984. "Jackson on Physical Information and Qualia". *Philosophical Quarterly* 34 (April): 147-183.

Husserl, Edmund. 1970. *The Crisis of European Sciences and Transcendental Phenomenology; an Introduction to Phenomenological Philosophy*. Evanston: Northwestern University Press, orig. 1936.

Indo, Yasuhiro. 2010. "Nerve growth factor, pain, itch and inflammation: lessons from congenital insensitivity to pain with anhidrosis". In *Expert Rev Neurother*. Nov;10(11):1707-24.

Irvine, Elizabeth. 2013. *Consciousness as a Scientific Concept: a Philosophy of Science Perspective*. Vol. 5. Dordrecht, New York: Springer.

Jackson, Frank. 1982. "Epiphenomenal Qualia". In *Philosophical Quarterly* 32 (April): 127-136.

Jackson, Frank. 1986. "What Mary Didn't Know". In *The Journal of Philosophy* 83, 5 (May), 291-295.

Kripke, Saul A. 1980. *Naming and Necessity.* Harvard University Press.

Levine, Joseph. 1983. “Materialism and Qualia: The Explanatory Gap”. In *Pacific Philosophical Quarterly* 64 (October): 354-361.

Levy, Neil. 2006. *Neuroethics: Challenges for the 21st Century.* Cambridge: Cambridge University Press.

Nagel, Thomas. 1974. “What is it Like to be a Bat?” In *Philosophical Review* 83 (October): 435-450.

Nida-Rümelin, Martine. 2015. “Qualia: The Knowledge Argument”. In *The Stanford Encyclopedia of Philosophy* (Summer Edition), Edward N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/sum2015/entries/qualia-knowledge/>.

Northoff, Georg. 2004. *Philosophy of the Brain*. Amsterdam Philadelphia: John Benjamins Publishing Company.

Northoff, Georg and Alexander Heinzel. 2006. “First-Person Neuroscience: A New Methodological Approach for Linking Mental and Neuronal States”. In *Philosophy, Ethics, and Humanities in Medicine* 1 (1): 3.

Perry, John. 2001. *Knowledge, Possibility and Consciousness*, Cambridge, MA: MIT Press.

Reardon, Sara. *A giant neuron found wrapped around entire mouse brain*. In *Nature News*. Nature Publishing Group, Feb 24, 2017.

Reisman, David. 2007. *Sartre's Phenomenology*. London; New York: Continuum.

Rowlands, Mark. 2001. *The nature of consciousness*. Cambridge: Cambridge University Press.

Sartre, Paul-Jean. 2013. *Being and Nothingness: An Essay on Phenomenological Ontology*, Routledge, orig. 1943.

Searle, John R. 1992. *The Rediscovery of the Mind*. Cambridge, MA: The MIT press.

Zanin, Fabio. 2011. *La Scomparsa Della Filosofia Naturale: Alle Origini Della Fisica Matematica*. Padova: Cleup.

Medline trend for “Consciousness”, URL = <http://dan.corlan.net/cgi-bin/medline-trend?Q=Consciousness> [Accessed September 2017]