DIALOGO - International Journal of the Proceedings of the Conferences on the Dialogue between Science and Theology

Volume 3 | Issue 2 Article 11

2017

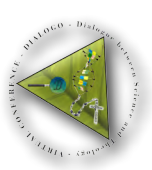
Non-locality of the phenomenon of consciousness according to Roger Penrose

Rubén Herce
University of Navarra, Spain rherce@unav.es

Recommended Citation
Herce, Rubén, "Non-locality of the phenomenon of consciousness according to Roger Penrose," Proceedings DIALOGO (DIALOGO-CONF 2017 SSC), DOI: 10.18638/dialogo.2017.3.2.11, ISBN: 978-80-554-1338-9 ISSN: 2393-1744, vol. 3, issue 2, pp. 127-134, 2017
Available at: http://www.dx.doi.org/10.18638/dialogo.2017.3.2.11

DIALOGO
CONFERENCES & JOURNAL
journal homepage: http://dialogo-conf.com

This paper was presented in the
Bridging Science and Religion together: SELF - SOUL - CONSCIOUSNESS. (DIALOGO-CONF 2017 SSC)
held online, on the Journal's website,
from May 19 - 26, 2017



Non-locality of the phenomenon of consciousness according to Roger Penrose

Rubén HERCE, Ph.D.
Assistant Professor of Philosophy of Science
University of Navarra
Pamplona Area, SPAIN
rherce@unav.es

ARTICLE INFO

Article history:
Received 24 April 2017
Received in revised form 4 May
Accepted 10 May 2017
Available online 30 May 2017
doi: 10.18638/dialogo.2017.3.2.11

Keywords:

Consciousness; Roger Penrose; Non-locality; Computation;

ABSTRACT

Roger Penrose is known for his proposals, in collaboration with Stuart Hameroff, for quantum action in the brain. These proposals, which are still recent, have a prior, less known basis, which will be studied in the following work. First, the paper situates the framework from which a mathematical physicist like Penrose proposes to speak about consciousness. Then it shows how he understands the possible relationships between computation and consciousness and what criticism from other authors he endorses, to conclude by explaining how he understands this relationship between consciousness and computation. Then, it focuses on the concept of non-locality so essential to his understanding of consciousness. With some examples, such as impossible objects or aperiodic tiling, the study addresses the concept of non-locality as Penrose understands it, and then shows how far he intends to arrive with that concept of non-locality. At all times the approach will be more philosophical than physical.

I. INTRODUCTION

Speaking of Roger Penrose and consciousness immediately refers to Stuart Hameroff, with whom he has written multiple articles (Hameroff and Penrose 2014a; 2014b). It is true that Penrose formulated its proposals more than two decades ago (Penrose 1996) and yet what is not so well known is the approach and motivations behind it. This paper proposes to travel back in time and recover the heuristic motivation behind some of Penrose's most recent proposals, bringing to light some interesting aspects for the debate on a consciousness that resists naturalization (Arana 2015).

In his essays, Roger Penrose makes an approximation to the mind-body relationship (Herce 2016). From the outset, he rejects a dualistic view of mind and body, as obeying different types of laws: physical on the one hand and free on the other. He

eISSN: 2393-1744, cdISSN: 2392-9928
printISSN: 2457-9297, ISSN-L 2392-9928
ISBN 978-80-554-1338-9

considers that what controls or describes the functioning of the mind must be an integral part of what governs the material properties of our universe (Penrose 1994, 213). In this sense, he is a naturalist and even a physicalist. However, according to Penrose, neither known physics nor computational activity would suffice to describe the functioning of the mind. There must be something else outside known physics that is *non-computational* in nature.

This article will begin by defining the different perspectives that Penrose observes regarding the possibility of artificially creating sentient beings. For this, the study will focus in particular on the first part of *Shadows of the Mind* (Penrose 1994), where Penrose delves into an argument that he had already exposed in *The Emperor's New Mind* (Penrose 1991).

The deepening of Penrose's arguments follows two paths: a negative critique, against those who think that our conscious mentality may be, in principle, fully conceived in terms of computational models; and a positive critique, to find out how and where this non-computational activity can be expressed. The first route is more rigorous than the second one. However, this article presents the reasons he argues for that positive quest for consciousness in the material realm, without focusing on the concrete solution, which is highly speculative, but extracting the heuristic motivations.

II. The framework for Penrose's stance

When Roger Penrose was asked in an interview what led him to cross the frontiers of physics and mathematics to investigate the phenomenon of consciousness, he answered:

It is a point of view that I formulated when I was in university in the 1950s. And I was fundamentally inspired by Gödel's theorem, which shows that mathematical truths cannot be reduced to calculations alone, and that to understand the mathematical realities we need to go beyond, out of mere computer rules. That is, no consistent system can be used to prove itself. What Gödel does is show how certain mathematical truths, that are beyond the reach of mathematical norms, can be established. So, the way we understand those rules allows us to transcend beyond the rules themselves. What that tells me is that our understanding is outside the norm. This is an aspect of the question that leads us to the next phase, our brain and the ability to think consciously, which is what separates us forever from computers: The most powerful and perfected of them can perform calculations of astonishing complexity with dizzying rapidity, but will never "understand" what it does. It is the result of how physical laws operate, and those physical laws have to be outside computational activity. Classical physics and quantum physics as we understand it today could be reduced to computation. So we have to go look beyond these two disciplines (…) What I speculate is that it is necessary to lay the foundations for the theoretical revolution that allows physics to include in its field the phenomenon of consciousness (Alfieri 2007, 126–27).

The above quote, although long, presents Penrose's compression frame for studying his proposal in relation to the phenomenon of consciousness. This proposal has the following starting point: a mathematician can understand mathematical issues, which are outside the norms that regulate these same mathematics. In such a way that mathematics cannot justify itself internally, but require an external justification. This idea connects with Gödel's incompleteness theorems and entails a two-level distinction between a level of conscious understanding of reality and a level that is not self-aware.

According to Penrose's position, there exists a physical world governed by precise laws, physical and mathematical, partly unknown. It is a predictable and calculable world, which is perhaps deterministic and also computable. In addition, there is another world related to consciousness. In this second one, which is not computable, is where some terms like soul, spirit, art or religion make sense (Penrose 1999, 82–84).

In turn, consciousness would have two areas of manifestation: a passive and an active one. The passive field would have to do with knowledge in the broad sense, and the active realm is associated with freedom and will. Penrose uses two terms "awareness" and "consciousness" to refer to the phenomenon of consciousness and, although he does not define them, he tries to clarify the terminology. He maintains that his position coincides with the common intuitive perception of the meaning of these concepts. In his scheme: "(a) 'intelligence' requires 'understanding' and (b) 'understanding' requires 'awareness'". In addition, 'awareness' would be the passive aspect of the phenomenon of 'consciousness', whereas 'free will' would be the active aspect (Penrose 1994, 37–40)

Penrose gives no further explanation, partly because he does not consider himself capable of philosophical precisions and partly because he conforms to common sense meanings. For his argument, it is enough to consider (1) that in order to understand it is necessary to be conscious and (2) that consciousness is a non-computable reality.

III. Four perspectives on the conscience-computer relationship

Penrose groups several arguments about the relationship between conscious thinking and computation in four perspectives:

A. All conscious thinking is computation. Just by performing the right computations, consciousness will be evoked.

B. Consciousness is a characteristic of the physical action of the brain. Any physical action can be simulated computationally, but the simulation itself cannot evoke consciousness.

C. Adequate physical action in the brain evokes consciousness, but this physical activity cannot be properly simulated.

D. Consciousness cannot be explained by physics, computation, or any other science.

None of these four types of relationships between conscious thinking and computation would be exclusive. Moreover, most authors would adopt more flexible positions. But the goal of Penrose is not to analyze all the possibilities but the most paradigmatic. Therefore, he focuses on these four positions, which associates to the approaches of four authors: Turing, Searle, Penrose, and Gödel respectively. And he submits these positions to four criticisms he calls: Searle's argument (against stance *A*), Chalmers's argument (against stance *B*), Turing's test or "scientific" argument (against stances *B* and *D*) and Gödel's argument (against positions *A* and *B*).

IV. Three critics to the above-mentioned perspectives

A. John Searle's argument

Perspective *A* would correspond with strong Artificial Intelligence and would be defended by Turing. According to this position, mental activity is simply the correct realization of a sequence of well-defined operations, such as those performed by any device with a simple algorithm. In this way, a well-programmed computer (or its programs) could understand language and

would have other mental capacities similar to human beings, whose abilities imitate. According to strong AI, a computer can play chess intelligently, make a smart move, or understand language. Similarly, the mind would have an extremely sophisticated algorithm, executed with exquisite subtlety, but nothing more. Therefore, any computer that possesses such an algorithm would be aware.

However, according to Penrose, the process of understanding is much richer than an algorithm that gives the right answer. Against the strong AI, it is directed the famous Searle's Chinese room argument (Searle 1980). This argument proposes a mental experiment by reduction to absurdity, whose central element is a human being performing an imaginary simulation of what a computer does. The human being inside a room follows instructions to order and handle Chinese symbols, although he does not know their meaning, much as a computer follows the algorithmic instructions of a program. Thus, as long as the human being manipulates the Chinese symbols following the instructions, it may seem that he understands Chinese, but he does not really understand anything. All it does is manipulate symbols without understanding the syntax or language (Cole 2015).

Another way of presenting this argument is as Penrose does. He presents a room that encloses a person without knowledge of Chinese but with the grammatical rules of the language and a perfect mastery of them. Later, this person is asked questions in Chinese whose meaning he does not understand, but to which he can give an adequate answer with the help of the rules. In this case, this person could respond well but would still not understand what he has answered.

Searle used this argument to criticize strong AI, while advocating a weak Artificial Intelligence, according to which brains would be equivalent to thinking machines. For Searle all aspects of understanding could be simulated, but simulation itself would not involve understanding. Therefore, for weak AI, computers would be a useful element for areas such as psychology or linguistics, because they could simulate mental abilities, but that would not mean that computers were intelligent.

Against the weak AI defended by Searle, Penrose presents a criticism by David Chalmers.

B. David Chalmers' argument

Penrose's B stance, which approximates the classical version of weak AI, holds that brain's actions could be simulated computationally. Even so, a similar external behavior would not be enough to know what the computer understands or feels and, therefore, to know if it is conscious, because consciousness, according to Searle, would be in what it feels and not in how it acts. Acting as a conscious subject would not be enough to ensure that you are aware. Therefore, the presence of consciousness would not be objectively discernible.

According to Penrose, this position has been criticized by David Chalmers (1996) in an argument that is directed only against stance *B* (weak IA) and leaves intact the rest of stances. The argument comes from the assumption, which Searle would accept, that in a human brain each of its neurons could be replaced in the future by a chip that works exactly the same. If this change were made individually, with each new replacement of a neuron, the person's inner experiences should remain unchanged. There would not be a nth neuron whose replacement would cause the loss of consciousness. Therefore, Penrose concludes with Chalmers, neither stance *B* is correct.

In summary and once these two

criticisms are considered separately, if Chalmers' argument and the Chinese room argument are combined, it turns out that Artificial Intelligence would be excluded as a whole, both in the strong version (stance *A*) and in the weak version (stance *B*). Artificial Intelligence would then be insufficient to explain the phenomenon of consciousness. Penrose comes to this same conclusion with his own argument. An argument known as Penrose's New Argument, perhaps because it constitutes a new deepening in the arguments Gödel-type that John Lucas had developed:

"I believe that our positions are very broadly in agreement, although the emphasis that I am placing on the role of the Gödelian argument may be a little different from his [Lucas]" (Penrose 1997, 7).

Since this paper is more interested in Penrose's proposal than in Penrose's criticism, it does not stop to analyze Penrose's New Argument (Lindström 2001; Herce 2014, 138–49), although he pretends to be more consistent and complete than the arguments by Searle and Chalmers (Penrose 1997, 9); and continues to expound the last of the arguments he poses against stance *D*.

C. The "scientific" argument

According to stance *D* - like *B* - the presence of a consciousness could not be detected scientifically, because it would not have experimentally verifiable manifestations. What differentiates the position *D* from the *B* is that according to the first the behavior of a human mind could not be simulated computationally, while for the second it would be possible to simulate. Although that does not mean the presence of a consciousness.

Penrose agrees with Mentalism - as he calls position D - in its claim that human mind cannot be simulated, but rejects it by holding that it is not scientifically possible to detect whether a being is conscious or not. He argues that consciousness can be detected scientifically, similarly to how the Turing test works.

This test is a test proposed in 1950 by Alan Turing to discover the existence of intelligence in a machine (Oppy and Dowe 2016). From a stance type A (strong IA), he assumes that if a machine acts in all respects as intelligent then it is intelligent. During the Turing test, a researcher in a room asks questions to a machine and a human being located in different rooms. His aim is to discover who the human being is and who the machine is, even though both can lie to him. Turing's thesis is that if the player and the machine are sufficiently skilled the researcher cannot distinguish who is who.

Penrose endorses the Turing test and generalizes it to what he calls the "scientific" argument. According to his position, it would be possible to detect the presence of a consciousness by means of scientific methods. He argues that the phenomenon of consciousness is not alien to scientific activity although is difficult to explain within the current scientific knowledge. He rejects the mentalist position because it is not scientifically testable, and because the enigma of consciousness already contains enough mystery without seeking solutions outside of science.

However, Penrose does not realize that the argument he uses to reject Mentalism can be used against the mathematical Platonism he advocates. There is enough mystery in the relationship between mathematics and physics without adding a Platonic mathematical world that is not scientifically testable either. In this sense, his critique of Mentalism is not consistent with his well-known Platonic stance in mathematics.

V. Penrose's proposal on computation and consciousness

As a scientific alternative to Mentalism, Penrose holds the position *C*. According to this perspective, computers never effectively simulate the conscious behavior of a human being. That is, there will always be someone during a Turing test who realizes that the computer does not understand.

"But viewpoint *C*, on the other hand, would not even admit that a fully effective simulation of a conscious person could ever be achieved merely by a computer-controlled robot. Thus, according to *C*, the robot's actual lack of consciousness ought ultimately to reveal itself, after a sufficiently long interrogation" (Penrose 1994, 14–15).

In fact, during the last decades an annual competition between computer programs that follows the standard established in the Turing test has been developing. However, so far, no program has managed to win the gold medal of the Loebner Prize, which is given to the couple (human-computer) that can deceive the judge.

Therefore and in other words, according to Penrose, (1) no unconscious object could be passed as a conscious subject. But (2) the presence of a conscious being would be scientifically detectable.

In studying the relations between consciousness and computability, Penrose adopts a posture that he calls *C*, more specifically *strong C*. According to this position, adequate physical action in the brains would be able to evoke consciousness. However, this action cannot be simulated by a computer: "We need a new physics that is relevant to brain activity." (Penrose 1999, 85) He thus departs from the stance *weak C*, for which the reason for the impossibility of such a simulation could be due to the non-computability of pure random phenomena observed in quantum mechanics.

VI. Non-locality of consciousness

Penrose's arguments on this issue are highly speculative. This is his most criticized and rejected hypothesis. In addition, it is an argument full of explanatory leaps. In his last three essays, he has not addressed this issue and between his first and second book has changed his thesis. However, they have an intrinsic heuristic value, especially if one simply analyzes the idea behind their proposal and not so much where the "conscious action" takes place[1].

For a compression of the scheme and as a background, it is necessary to resort to the concept of *non-locality* given in the *aperiodic tiling problem* and in Penrose's impossible objects, two key elements of his work (Herce 2014, 25–30). According to this concept there may be a level of determination that is above the local level. That is, what locally seems indeterminate, from a higher level could be determined, such as in the Penrose's triangle or staircase. These objects, seen partially (locally) are possible, but seen together (non-locally) are impossible.

Similarly, in the aperiodic tiling problem,

1 Penrose associates the phenomenon of consciousness with a coordinated action on a large number of brain neurons that would be caused by the orchestrated reduction of the state vector (Orch OR) in the neural microtubules. Since microtubules are found in many structures of living things, Penrose goes on to argue that even the paramecia would perform some conscious activity, because they have a cytoskeleton formed by microtubules. Depending on the complexity of the structures and the amount of microtubules involved in each orchestrated reduction, there would be degrees of consciousness, higher in mammals, and very special in the case of human beings. Thus, for Penrose consciousness emerges from the material and he points to microtubules as structures in which the conditions of possibility of conscious actions could be given due to the effects of quantum gravity during the collapse of the wave function. As it has been said: highly speculative.

when one wants to cover a surface non-periodically with a finite number of tiles, no pattern of repetition is found. However, from a non-local level that type of tiling can be found; and, in fact, following a very simple scheme with only two types of tiles, it has a deterministic and non-computable evolution. Robert Berger showed that the evolution of this scheme cannot be simulated by any computer, because there is no algorithm capable of deciding whether a finite set of tiles will cover a surface (Berger 1966). This scheme is then governed by non-local rules that are beyond computation.

Therefore, according to the Penrose scheme there are two levels: a non-computable upper level that influences the lower level. In analogy to how the aperiodicity of the aperiodic tiling (from the upper level) influences the lower level, without being locally detectable.

At the local level, everything might seem determined and computable. Only when viewed from a higher level does the noncomputability appear. So also, understanding, knowledge and consciousness would be given at a higher level that is not computable.

These considerations of Penrose give rise to consider that there could be types of higher order *non-computability* (Penrose 1999, 100) involved, for example, in the way the universe evolves or in human freedom. He defends thus the existence of several levels, not only of two, each of which could have the characteristics of a "determinism not computable" with respect to the superior level.

From here, a couple of points for further research should be highlighted. The first is the recursive aspect of many physical phenomena and their possible relation to consciousness, as some authors have explored (Hofstadter 2013). And the second is the deduced conclusion that consciousness is situated on a higher level than the physical, although it could emerge from it. Penrose, does not attempt to include consciousness on the same level of physical or mathematical causes, nor to separate it completely from them. He thus leaves the way open to a consciousness that interacts with physical levels, although the way in which that relationship takes place remains the greatest mystery.

Conclusions

This work has started by pointing out the different positions that Penrose distinguishes in relation to whether or not the consciousness is computable. From there, it has shown the criticisms made by Penrose and with what position he stays. Having defined his position as strong C, the paper has explored what such a position consists of and how Penrose's comprehension of consciousness revolves around the concept of non-locality, which has also been explained. From this last idea and taking one more step, this paper concludes saying the following.

From a local point of view, Penrose points out the existence of determinate and computable realities that coexist with others that seem indeterminate, because in them occur decisions or novelties that are not computable. However, the decision or novelty that appears on a certain level could be determined by some law of a higher level.

Such a law would, for example, be responsible for preventing Penrose stairs from actually existing, albeit locally seemingly possible, or allowing aperiodic quasi-crystals to be actual physical configurations, although they do not have a local pattern that structures them. Therefore, Penrose deduces, that at the local level everything would be determined but not everything would be computable. The apparent indeterminacy of the non-computable realities would be determined

from the upper or global level. In short, the non-computability manifests locally, but refers to a non-local element.

A further step, given by Penrose, is to admit the possibility that there are several levels of determination. Thus, on the higher level you could find both the law that governs the universe and the consciousness that acts freely. From these two levels of universal law and personal liberty, events at lower levels would be determined.

This position of Penrose tries to maintain an equilibrium, which hardly prevents to end in one of the two previously rejected ends: either a materialism where freedom is only apparent, because everything is determined by a higher law, or a scientifically indemonstrable mentalist dualism that gives room for freedom. It is not clear where it ends.

Bibliography

[1] Alfieri, Carlos. "Roger Penrose : 'Creo En Un Universo de Ciclos Sucesivos.'" *Revista de Occidente*. 2007

[2] Arana, Juan. *La Conciencia Inexplicada : Ensayo Sobre Los Límites de La Comprensión Naturalista de La Mente.* Madrid: Biblioteca Nueva. 2015

[3] Berger, Robert. "The Undecidability of the Domino Problem." *Memoirs of the American Mathematical Society*, no. 66/1966: 1–72. doi:10.1090/memo/0066.

[4] Chalmers, David John. *The Conscious Mind : In Search of a Fundamental Theory.* Philosophy of Mind Series. New York: Oxford University Press. 1996

[5] Cole, David. "The Chinese Room Argument." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Winter 201. Metaphysics Research Lab, Stanford University. 2015

[6] Hameroff, Stuart, and Roger Penrose.. "Consciousness in the Universe." *Physics of Life Reviews VO - 11*, no. 1/2014a. Elsevier B.V.: 39. doi:10.1016/j.plrev.2013.08.002.

[7] ———. 2014b. "Reply to Criticism of the 'Orch OR Qubit' – 'Orchestrated Objective Reduction' Is Scientifically Justified." *Physics of Life Reviews* 11 (1): 104–12. doi:10.1016/j.plrev.2013.11.014.

[8] Herce, Rubén. De La Física a La Mente : El Proyecto Filosófico de Roger Penrose. Madrid: Biblioteca Nueva. 2014

[9] ———.. "Penrose on What Scientists Know." *Foundations of Science* 21 (4)/2016: 679–94. doi:10.1007/s10699-015-9432-0.

[10] Hofstadter, Douglas R.. *I Am a Strange Loop*. Basic books. 2013

[11] Lindström, Per. "Penrose's New Argument." *Journal of Philosophical Logic* 30 (3)/2001: 241–50. doi:10.1023/A:1017595530503.

[12] Oppy, Graham, and David Dowe. 2016. "The Turing Test." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Spring 201. Metaphysics Research Lab, Stanford University.

[13] Penrose, Roger. *The Emperor's New Mind : Concerning Computers, Minds and the Laws of Physics / Roger Penrose ; Forew. by Martin Gardner.* New York: Penguin Books. 1991

[14] ———. 1994. *Shadows of the Mind : A Search for the Missing Science of Consciousness.* Oxford [etc.] : Oxford University Press, 1994.

[15] ———. 1996. "On Gravity's Role in Quantum State Reduction." *General Relativity and Gravitation* 28 (5): 581–600. doi:10.1007/BF02105068.

[16] ———. 1997. "On Understanding Understanding." *International Studies in the Philosophy of Science* 11 (1). Routledge: 7–20. doi:10.1080/02698599708573547.

[17] ———. 1999. Lo Grande, Lo Pequeño Y La Mente Humana. Cambridge University Press.

[18] Searle, John R.. "Minds, Brains , and Programs." *Behavioral and Brain Sciences* 3 (3)/1980: 417–57. doi:10.1017/S0140525X00005756.