Proceedings of the International Conference on Social Psychology and Humanity Studies
DOI: 10.54254/2753-7064/6/20230083

The Possibility of Artificial Qualia

Xiang Wang[1,a,*]

[1] *College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China*
a. wangxiang-@zju.edu.cn
**corresponding author*

Abstract: The discussions about qualia are mind-body problems. If artificial intelligence provides a functionally identical body, will it possess the same conscious experience? The possibility of artificial qualia, which is on the cutting edge of the development of artificial intelligence, will be the main topic of this essay. This paper argues the presupposition of the possibility of artificial qualia is the possibility of strong AI. Searl was the first person to question the possibility of strong AI by formulating the Chinese Room Argument. This paper introduces common objections from artificial intelligence practitioners and the development of modern AI, proving the invalid of the objection and claiming the possibility of strong AI. Next, this paper follows David Chalmers' argument about the principle of organizational invariance and reconstructs Chalmers' two Reductio ad Absurdum arguments to rebut objections. By formulating fading qualia argument and dancing qualia argument, the absent qualia argument and the inverted qualia argument are proved wrong. The principle can support the possibility of artificial qualia.

Keywords: qualia, strong AI, the principle of organizational invariance

1. Introduction

Are artificial persons possible? What is an artificial person? What are the fundamental elements needed to become a person? Is consciousness or perception necessary to be a person? Moreover, can artificial intelligence be referred to as an artificial person if it possesses those qualities? Is that experimental phenomenon essential to completing humans? Those experimental phenomena are qualia. The paper will focus on the possibility of artificial qualia, and the author will argue that artificial qualia are possible.

Qualia are the experiences humans possess. C. I. Lewis introduced the term "qualia" in 1929 [1]. Qualia include sensory experiences: the experience of seeing red (visual sense), the experience of hearing a musical tone (auditory sense), and the experience of tasting a piece of chocolate (taste sense). Qualia also include the feeling of being in pain and emotions like depression. Qualia are private first-person and phenomenal consciousness. Qualia are the core issues in discussions of the philosophy of mind. Traditionally, qualia are intrinsic characteristics of experience that can be directly accessed through introspection [2]. Modern theories concentrate on the types of mental states that have qualia, whether qualia are intrinsic qualities of those who experience them, and the relationship between qualia and the physical world, both within and outside the head [3].

Qualia are the phenomenal consciousness humans possess. If the usage of the term is limited to humans with the law of nature, any artificial properties could not be possible. If the person only refers

DOI: 10.54254/2753-7064/6/20230083

to natural human beings, an "artificial person" could not be possible by nature. However, in most cases discussing the artificial person, the expression needs to be interpreted in a broader sense. The artificial person has all the qualities humans possess. Similarly, artificial qualia have all the qualities humans possess. They do not necessarily require physical identities like neurons or brains.

The possibility of artificial qualia is a fundamental question because it could imply the edge of the development of AI, namely, estimate how far AI can achieve. The following section will prove a presupposition for the thesis. By responding to some objections to the possibility of strong AI, I argue that strong AI is possible. Section three will reconstruct Chalmer's vindication of the principle of organizational invariance. By proving the falsity of the absent qualia argument and the inverted qualia argument, the principle of organizational invariance proves the possibility of artificial qualia. Actually, the question of robots and artificial intelligence with consciousness also involves other disciplines like psychology, brain science, and computer science. However, the paper will not cover them but instead focus on the metaphysical existence of artificial qualia.

2. Presupposition of the Possibility of Strong AI

Artificial Intelligence is generally divided into Strong AI and Weak AI. The difference between the two lies in their goals: Strong AI aims to create artificial persons that are fully human in every way and possess the entire range of human's mental faculties [4], including phenomenal consciousness; Weak AI is building machines that can perform some of the functions of human beings [5]. Weak AI is not "weak". Weak AI could also pass Turing Test and the Total Turing Test [6]. Those tests are designed to determine whether a machine is truly intelligent [7].

Weak AI is generally considered not to have qualia when considering the issue of qualia. Strong AI by definition should have qualia if it needs to achieve the same function or perception as a human [4]. However, the question is whether the existence of Strong AI is possible, that is, whether the artificial person is possible. There is a basic consensus on the existence of Weak AI, as it is widely acknowledged that it is difficult to overthrow Weak AI [8]. On the contrary, Strong AI is different, for the term was introduced by Searl to question its existence [9].

The artificial qualia could be considered as one property of the artificial person. And the possibility of an artificial person equals the possibility of Strong AI. According to the form of the categorical syllogism, if the existence of Strong AI is impossible, then the existence of artificial qualia is also impossible. Therefore, the first question to answer regarding the possibility of qualia is whether Strong AI exists.

2.1. Objections to Strong AI

The Chinses Room Argument (CRA), developed by John Searle in 1980, is the most well-known theory that challenges the viability of strong AI. Searle proposes a thought experiment that he is inside a room and does not know any Chinese. People outside the room send cards with questions in Chinese through a slot. A symbol processing program is written in English (Searle speaks English). Searle follows the instructions for manipulating Chinese symbols. Searle aims to point out that the operational process of a computer is just what Searle-in-box does: manipulate symbols based on their syntax. CRA shows that even though Searle-in-box passes the Turing Test, Searle-in-box does not understand any Chinese. Therefore, Strong AI is not possible.

J.R. Lucas [10] asserted that Gödel's first incompleteness theorem implies that no machines may ever achieve human-level intelligence, which is another effort to refute Strong AI. Many philosophers have come up with vindications based on Lucas' argument.

DOI: 10.54254/2753-7064/6/20230083

2.2. Defenses for Strong AI

Some people respond to CRA by claiming that no specific area of the human brain really understands English, either. The fact is that the brain as a whole system knows English. Likewise, the Chinese room as a whole system knows Chinese; the person inside the Chinese room as the component of the system does not know Chinese. Therefore, the possibility of Strong AI still exists.

In support of Strong AI, Borge [11] claims that the implementation of the computer program will make a system with built-in intentionality possible. Moreover, Borge formulates the "Model Address Argument" to challenge Searle's claim and to prove the potential success of Strong AI. One of the most famous responses is Rapaport's. He asserts that although AI systems are syntactic, the proper syntax might be semantics.

In the last decade, many advanced technology companies today have created humanoid robots, such as Pepper from Softbank robotics, Sophia, and Amaka from Engineered Arts. They are very intelligent, such as some of them can generate natural conversations or recognize humans' emotions and respond to them. It is intuitively difficult to deny the possibility of Strong AI. For the earliest objection to Strong AI, CRA is widely opposed by AI practitioners like scientists and engineers. Advocates of strong AI claim that the CRA is "unsound" and "silly". The thought experiment of the Chinese room is far removed from the practice of AI. Those scientists and engineers think the development of more sophisticated robots will silence CRA [5]. Considering the development of technology, it is more likely to accept the possibility of strong AI, or at least the possibility is imaginable. Based on this presupposition, the possibility of artificial qualia could be discussed. And the following question is on what principle the possibility of artificial qualia is based.

3. The Principle of Organizational Invariance Enables Artificial Qualia

This paper tries to prove the possibility of artificial qualia through the principle of organizational invariance. The exposition on the principle of organizational invariance regarding qualia was presented by David Chalmers [12].

One way to distinguish the theories of qualia is by dividing them into five categories: functionalism, physicalism, representationalism, eliminativism, and naturalistic dualism. Chalmers [12] defines this principle as nonreductive functionalism. And those two objections to the principle are typically objections to functionalism. General objections to functionalism have two ways: 1) Absent Qualia: one system has qualia while another does not, but they are functionally identical, and 2) Inverted Qualia: two systems have very different qualia, but they are functionally identical.

Chalmers proposes the principle of organizational invariance as: the conscious experience is constant, across systems with the same fine-grained functional organizations [12]. It means that any system with conscious experiences and the same fine-grained functional organization will have identical conscious experiences. The key factors of the functional organization are the number of components and their dependent relations and interactions. In the case of artificial qualia, the organization with the electronic chips formulating the artificial person could have the same experience as the neuron organization in the human brain. In this case, the artificial person experiences so-called qualia identical to human's, proving the possibility of artificial qualia.

Chalmers uses two Reductio ad Absurdum arguments to reject those two objections to functionalism (the Absent Qualia and the Inverted Qualia). Those who support the Absent Qualia argument claim that consciousness requires biological components [13]. Chalmers first supposes that absent qualia are true, then he proposes a thought experiment. He claims that if absent qualia are possible, then fading qualia are possible. Those who support the Inverted Qualia argument claim that the conscious experiences of robots are totally different from those of humans. Chalmers uses the form again: he supposes that inverted qualia are true, then he claims that if inverted qualia are true,

DOI: 10.54254/2753-7064/6/20230083

then dancing qualia are true. Chalmers proves the falsity of fading qualia and dancing qualia, so he proves the falsity of absent qualia and inverted qualia. Ultimately, Chalmers proves the truth of the principle of organizational invariance.

First, the fading qualia argument asks the audience to conceive of a scenario in which a little aspect of "me" (the functional organization) changes. A small amount of the neuron is replaced by silicon chips. The functional organization will not be influenced if the silicon chips function well. And the same process goes on, replacing more and more neurons with chips until there are no biochemical components in the original functional organization. According to the absent qualia argument that he supposed is true, the silicon systems do not have qualia in the last situation. Chalmers considers the spectrum of changes between the two ends and tries to find how qualia disappear in the process. And he gives two possibilities, one is fading qualia, and another is suddenly disappearing qualia. The second indicates inconsistency, so he thinks the antecedent plausibility is very low. The first argument, the fading qualia argument, is logically possible, but its experiences notions are totally wrong. Considering this contradiction, Chalmers proposes a more reasonable hypothesis that qualia do not fade. Therefore, the absent qualia argument is wrong, and the principle of organizational invariance is true.

Second, the dancing qualia argument is another Reductio ad Absurdum argument. Similarly, Chalmers first supposes that the inverted qualia are true. And Chalmers formulates another thought experiment that there is one silicon system that has inverted qualia with "me" (the original organization). And there is a switch to control those two organizations. After flipping the switch, the conscious experiences are redirected. It means that the qualia of the original organization are changed to the silicon system and vice versa. The case after flipping the switch is that the systems now have inverted qualia as they had before. The absurdity is that there is no possible way to make judgments of those inverted experiences, and so they should operate without any changes. Therefore, Chalmers proves that dancing qualia could not happen and inverted qualia are wrong.

After those two thought experiments with Reductio ad Absurdum arguments, Chalmers concludes that the most plausible theory is that qualia still exist when replacing neurons but preserving functional organization. It implies that the existence of qualia is only influenced by the functional organization. Therefore, Chalmers counters those two objections to the principle of organizational invariance. And he claims that the principle of organizational invariance is true and qualia (the conscious experience) are determined by the functional organization. And the principle of organizational invariance confirms that artificial functional organizations will have identical conscious experiences with humans, which means the principle confirms the possibility of artificial qualia.

However, according to van Heuveln et al., Chalmers' Dancing Qualia Argument is invalid [14]. They point out that by assuming the falsity of the intended conclusion, no untenable stance is obtained. Chalmers' dancing qualia argument rests on backgrounds assumptions: anti-materialism, and natural dualism. Van Heuveln et al. think that Chalmers bases his dancing qualia argument on these presumptions and seeks to prove the validity of organizational invariance. And they argue that Chalmers' argument is invalid, because, given those assumptions, assuming the falsity does not lead to a contradiction.

Moreover, although Chalmers proved the wrongness of the Absent Qualia Argument, however, in the case of the artificial person, this paper does not specify the constituent components of the artificial person. The position of this paper is that computing chips of metal or silicon can constitute an artificial person. However, it also retains the possibility of a version with biochemistry as an element to constitute an artificial person.

DOI: 10.54254/2753-7064/6/20230083

Chalmers responds to the objections of the principle of organizational invariance, which supports the goal of Strong AI. With the principle of organizational invariance, artificial systems could have qualia that are identical to humans and prove the possibility of artificial qualia.

4. Conclusion

This paper first discusses the definition of qualia and then clarifies two categories of artificial intelligence. They imply that the presupposition of discussing the problem of artificial qualia is the possible existence of Strong AI. To prove the possibility of strong AI, this paper responds to the most famous objection CRA made by Searl. This paper then points out counter-intuitiveness to deny the potential of strong AI, especially under the circumstances of the rapid development of technology.

The principle of organizational invariance proposed and vindicated through Chalmers' argument for Qualia Argument, proves the possibility of artificial qualia. The defense for the principle stands by proving that the two objections to the principle are wrong. The paper reconstructs the process of the two Reductio ad Absurdum arguments.

There are some objections to Chalmers' arguments as the paper mentioned in the previous section, but the paper fails to respond to those. Moreover, this paper talks only about metaphysical or logical possibilities. However, the creation of AI requires much more realistic factors. The paper does not include the gap between the theoretical possibilities and the limitations in the practical process.

References

[1] *Lewis, C.I. (1929). Mind and the World Order. New York: Charles Scribner's Sons.*

[2] *Kind, Amy. "Qualia," The Internet Encyclopedia of Philosophy, ISSN 2161-0002, https://iep.utm.edu/qualia/, Dec. 31, 2022.*

[3] *Tye, Michael, "Qualia," The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2021/entries/qualia/>.*

[4] *Searle, J. (1997). The Mystery of Consciousness, New York, NY: New York Review of Books.*

[5] *Bringsjord, Selmer and Naveen Sundar Govindarajulu, "Artificial Intelligence", The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/>.*

[6] *Harnad, S. (1991). Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem, Minds and Machines, 1.1, 43-54.*

[7] *Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59, 433-460.*

[8] *Bringsjord S. & Xiao, H. (2000). A Refutation of Penrose's Gödelian Case Against Artificial Intelligence. Journal of Experimental and Theoretical Artificial Intelligence, 12, 307-329.*

[9] *Searle, J. Minds. (1980). Brains and Programs, Behavioral and Brain Sciences, 3, 417-424.*

[10] *Lucas, J. R. (1964). Minds, Machines, and Gödel, in Minds and Machines, A. R. Anderson, ed., Prentice-Hall, NJ: Prentice-Hall, 43-59.*

[11] *Borge, Steffen. A Modal Defence of Strong AI. (2007). In Dermot Moran Stephen Voss (ed.), The Proceedings of the Twenty-First World Congress of Philosophy. The Philosophical Society of Turkey, 127-131.*

[12] *Chalmers, David J. (1995). Absent qualia, fading qualia, dancing qualia. In Thomas Metzinger (ed.), Conscious Experience. Ferdinand Schoningh, 309-328.*

[13] *Longinotti, D. (2018). Agency, Qualia and Life: Connecting Mind and Body Biologically. In V. C. Müller (Ed.), Philosophy and Theory of Artificial Intelligence Springer International Publishing, 44, 43-56.*

[14] *Van Heuveln, B., Dietrich, E., & Oshima, M. (1998). Let's Dance! The Equivocation in Chalmers' Dancing Qualia Argument. Minds and Machines, 8(2), 237-249.*