Features necessary for a self-conscious robot in the light of "Consciousness Explained" by Daniel Dennett

What are the features necessary for a self-conscious robot in the light of "Consciousness Explained" by Daniel Dennett?

Contents

The soul in a machine

Self-consciousness relates to important themes, such as sentience and personhood, and is often the cornerstone of moral theories (Warren, 1997). However, not much attention is given to future moral creatures of the earth: robots. This may be due to the unsettled status of their experience, which is why I have chosen to find the necessary features of self-consciousness in them. Philosophy of mind is also my interest which I have developed since I rejected the idea of souls and could not account for my own phenomenal experiences. Among the current answers to the hard problem of consciousness (Chalmers, 1995) one stood out to me – Daniel Dennett's multiple drafts model of consciousness presented in "Consciousness Explained" (Dennett, 1991) and his other works. Daniel Dennett is a self-described illusionist (Dennett, 2016) and a functionalist. His position is synonymous to holding self-conscious experiences, such as deliberating over one's pain or reflecting over one's actions insufficient to postulate existence of a metaphysical self. He believes instead that experiences are functional realizations which do not require any self, and the feeling of being conscious or self-conscious is actually a "user-illusion" (Dennett, 1991), a term describing the unjustified belief of being in control of one's actions. In his views, a selfhood as it is popularly conceived is contradictory. Therefore, he rejects realism about mental events, a position also called naïve realism (Siegel, 2011), on the basis that it is unverifiable and there are insufficient grounds to accept mental events (Dennett, 1988). He adopts a similar scepticism towards other theories of minds, hence nearing eliminative materialism in self-consciousness and related theme of pain (Dennett, 1978). His multiple drafts model of consciousness seeks to ground our beliefs about self-consciousness in the specific architecture of mind, which is particular to every species and kind of being. As such it is a particularly flexible theory that examines experiences of non-humans without falling into binary thinking – believing that a being has as rich inner life as humans do or none. This is why it is a great starting point for establishing necessary conditions for consciousness in robots. In this essay, I will often use words *computer, programme, robot* interchangeably for the purpose of clarity and take as the working definition "autonomous machines capable of sensing their environment, making decision and performing tasks similar to human actions". I chose to explore self-consciousness instead of committing to the issue of consciousness, because consciousness can be defined in a weak sense, in which even a thermostat is "conscious" of the temperature, and consciousness in a strong sense, which invokes the internalism and externalism debate that is less connected to the nature of experiences than self-consciousness. My exploration pertains first and foremost to the philosophy of mind, of which I will examine

the most relevant positions and decide to commit to functionalism (Block, 1980). Then I will advocate for explaining away of some of phenomena of self-consciousness and believing self-consciousness to be a sort of illusion. From there I will conclude that robots could fall for the same illusion of self-consciousness and therefore become self-conscious. I will argue, that self-consciousness can be established using a modified Turing Test, which would determine if a robot believed in a meaningful narrative about itself (Ricoeur, 1991) and hence be self-conscious. This will lead to answering the question of this essay: what are the features necessary for a self-conscious robot in the light of "Consciousness Explained" by Daniel Dennett?

The attractions of mind stuff

Consciousness has been a highly controversial matter at least since Descartes' time (Descartes, 1641). The philosopher proposed a distinction between res cogitas and res extensa, the former denoting mind and the latter the physical world. Consciousness and its phenomena – such as pain, smells and colours - could thus be explained as being immaterial, and therefore not explicable in terms of human anatomy, while the mechanistic aspects of human body were part of the world the mind was to control. This approach – known as Cartesian dualism – has been criticized because it was incompatible with our understanding of the physical world based on science. Barring that, it required that the material and the immaterial had to act in sync, perhaps by God's intervention as suggested by Malebranche (Watson, 2016) or earlier by Abu al-Hasan al-Ash'ari. Otherwise it was rejected by naturalists such as Hume on the grounds there is no such thing as a self (Hume, 1740). Furthermore, it left open the question of whether the immaterial could leave the body behind, which would nevertheless continue to behave as if it was conscious (Chalmers, 1996). We would call such a being a philosophical zombie – outwardly it would behave exactly like a body with a soul, but it would not experience any private sensations (Campbell, 1970). Dennett's contribution to the debate was suggesting that the problem of philosophical zombies and conscious experience are both doesn't have any signifier, and that they are a sort of fiction. Both concepts are fundamentally flawed because they include essential properties, like infallibility, privacy and intrinsity, that cannot co-exist, as in a well-documented phenomenon known as "reactive disassociation".

What is self-consciousness?

To find features necessary for a self-conscious robot, we have to have some theory of conscious experience. On one account, each mental state – quale - corresponds to a brain state. Pleasure

and pain are consequently synonymous to events in the body, such as release of specific neurotransmitters. This is what happens according to identity type theory (Schneider, 2009). Unfortunately, it tells us very little, because brains of animals are very different, and yet we know when one of them is in pain. But then what is pain? Pain is usually what links an injury and the belief that one is in pain along with a reaction. We could therefore call it a sort of a function (Block, 1980). This view – called functionalism – is what I endorse in the essay. The most common objection to it is that it does not seem to explain what some would call "actually feeling pain". Let's call sceptics of functionalism qualophiles. Qualophiles would argue, that we can imagine reacting to pain without experiencing any, and that even if pain was only an evolutionary adaptation, then we would still have to explain why such a mechanism would come with phenomenal experiences. If the phenomenal experiences were part of the mechanism, they say, people could at most be philosophical zombies, but they could not genuinely feel anything, because the justified true belief that one is in pain requires the existence of phenomenological experiences, which they don't have. Qualophiles would rather have qualia stand for conscious experiences. Thus, qualia would appear not to be part of the mechanism of behaviour, but rather additions to the world[1]. This raises several questions. For one, if qualia were not parts of the mechanism, we could not tell if they existed or not nor could we claim to actually feel anything because qualia would not influence our actions and therefore our judgements about them. We could have beliefs about whether we are experiencing anything, but such beliefs could never be said to be justified. We would therefore have no good reason to believe in them unless we proposed some special relation to the phenomenal states that could sidestep the causality of the universe. This is the solution offered by naturalistic dualists such as David Chalmers (Chalmers, 1996). However, it meets the same objections as Cartesian theatre does (Bayne, 2001).

Perhaps we are trying to draw up a catch-all conjecture that explains all the supposed phenomena of consciousness, but when the analogy fails, we naturally do not want to deny the existence of experiences, so we end up believing unjustified theories. The existence of qualia is up to debate, and we might find a better way to explain the phenomena of experience and find the necessary features for self-consciousness in robots.

Fundamentally we can approach the problem of consciousness in two ways: consciousness can either be explained, just as a neuroscientist can show how the eye turns different wavelengths

[1] The view that mental events have no effect upon physical events is called epiphenomenalism, (Robinson, 1999).

into electrical signals, or explained away by demonstrating why it seems like it exists[2]. We might find the latter approach better in situations when what is actually happening differs from what it seems like – for instance, watching a magician split a person in half we will not assume anyone actually gets hurt, but rather try to explain why it would look like it from our perspective. This is the reason I picked Daniel Dennett's "Consciousness explained" – its goal from the beginning is to account for why intrinsic qualities seem to exist, but and the same time they are not reconciled with neither philosophy nor science. In the spirit of the text, we might find self-consciousness to be a sort of illusion. After the magic tricks employed in this illusion are analysed and demystified, it will be a matter of finding out how could robots become convinced of their self-consciousness. This will be synonymous to answering the research question of which features of consciousness are necessary for a self-conscious robot.

How to imagine a self-conscious robot?

To explain away self-consciousness, we may find it useful to divide our attitudes towards different types of objects in a three-level model of abstraction as proposed by Dennett (Dennett, 1987). The most fundamental level is the physical stance, which describes a system on the level of atoms and particles. It can be useful in predicting the behaviour of a rock, such as when describing its fall to the ground. It would be impractical, however, to consider in this way more sophisticated mechanisms, like thermostats. A more useful approach towards such items is a design stance, which treats them as purposeful tools. Looking through this lens, we sacrifice the precision of the physical stance for predictability in a range of "normal" situations, such as all those where the temperature does not cause the thermostat to fail. The third level is the intentional stance, which presupposes the reason and agency of a being, allowing us to speak of beliefs, desires (in the narrow sense) of animals, artificial intelligence and humans. Adopting an intentional stance towards a being suggests that the subject is self-conscious, as it exhibits goals of its own and awareness of its existence. This three-level approach is present in much of our thinking, and will become helpful in overstepping the metaphysical divide between humans and robots, because it allows us to think of computers as prone to the same illusion of consciousness without fallacious comparison to humans. We can, for instance, think about Stockfish 14 chess algorithm as having a goal. It was designed to win, and to predict its behaviour it would be more practical to find the best moves it can make instead of analysing its physical or design stance. Playing against the algorithm we might find it easiest to ascribe it

[2] The argument in its entirety is presented in (Dennet, 2016).

possession of information (the knowledge of position of chess pieces) and having goals (trying to win) (Dennett, 1987). It is only a small step away from speaking of beliefs and intentions of the machine, although those might be more relevant when considering more complex instances of artificial intelligence. This is still, however, a useful analogy and not an actual ascription of any of such states; this is because possession of mental states, goals, beliefs, desires has a more specific meaning which only applies to humans. Dennett's approach is this respect is quietist, as he mostly focuses on defending the possibility of consciousness in robots, but hardly asserts anything about their inner experiences. An underlining theme of Dennett's argument is verificationism (Creath, 2011) – it is the reason for constant scepticism about assertions about qualia. This is why in what way does a computer really have beliefs is irrelevant to his theory and the perspective of intentional stance. My goal is to assert self-consciousness by finding how robots could fall under the same illusion of self-conscious.

How can humans actually be conscious is unimportant once we are able to explain why they would be compelled to think they are, and once robots appear to be equally convinced about their experiences we will have to consider what makes our judgements about experience true. I will rest the truthfulness of statements about experience on the narrative of a given being, for neither experiences nor illusion of experiences can verify statements. Proving that a person feels pain cannot be justified by experiences, because existence of experiences is informed by experiences. Similarly, claiming that a person falls illusion to being in pain does not explain why would such a person believe that. Therefore, both explanations are to a degree circular and we have to look for another criterion for truthfulness of statements asserting experience. I believe it to be narrative. I do not assume existence of any artefacts of self-consciousness, but merely search for a way in which we become convinced of it and speculate how robots could as well.

To determine whether robots could believe they were self-conscious, however, we also have to be more exact on what "consciousness" is. In the day-to-day sense it can mean anything from the state of being awake to experiencing the burden of postmodern condition. In philosophical writing it is often defined in such a way as to cohere with any view of confirming realism of experiences, such as epistemological subjectivism of John Searle (Searle, 1997) or David Chalmers special acquaintance with conscious experiences (Chalmers, 1996). Therefore, the conclusion – that phenomenal experiences exist – is already implicit in the premise (Dennett, 1988). Instead, the definition should account for how a being can become convinced of having inner experiences. The working definition in this paper will therefore be "a state of being aware

of one's experiences". Reporting on an experience will often be a mark of being conscious of it, but it does not necessarily follow from it. We know, for instance, that a computer does not have the faculty to experience pain, and so a programme reporting "I feel pain" will be dismissed. However, if a computer has means to make a judgement on some matter, by some Bayesian metrics we can treat its report as a justified belief. Having a command typed on a keyboard could equivalently base the claim that the computer is conscious of the input.

Setting a criterion for truthfulness of a statement about experience precludes us from mistakenly adopting intentional stance towards an object and taking it to be conscious. If we take the statements produced by intelligent chatbots to be expressions of belief, we seem to be offering the status of conscious beings even to the supposedly mindless artefacts. But is that so absurd? Charles Darwin spent several years studying the behaviour of earthworms, which he regarded as possessing some sort of sentience and even reason (Darwin, 1881). Complexity of chatbots largely exceeds that of earthworms, but consciousness of both is limited by what they can be conscious of. Neither computers nor earthworms have yet become aware of the burden of the postmodern condition. If they did, would there be anything stopping them from being conscious in the same sense humans are? Could they become self-conscious? How would we know?

Modified Turing Test

How do we know if something is conscious in the first place? We can again choose between two approaches. Either we think that consciousness exists in some real way, for instance qualia, or we try to explain why we would come to think that something is conscious. To do that, we may use the intentional stance. Children learn to adopt an intentional stance towards others during infancy (Öner, 2010), which coincides with development of language. They learn it solely by interaction with others. Some children believe their toys also to be conscious but shun it when they get older. The reason might be that they are able to interact with and get response from other humans, but not so much with toys. Robots would probably land in the middle of the gap between humans and toys and so it would be difficult to determine if they would be considered conscious or not. Talking robots would be more convincing than those solely performing a task, but the most convincing ones would probably be able to both conversate and be able to interact with the real world. This would require them to possess some degree of intentionality[3] (Pierre, 2003), that is the ability of mental states to be about external properties and

[3] Intentionality in this is different from the meaning of *intentional stance*. Intentionality are used in both contexts throughout the essay.

states of affairs. This does not seem out of reach for at least some sorts of artificial intelligence (Mason, 2017). Unfortunately, modern robots are usually very limited, as they possess either apparently excellent linguistic capabilities or the tools to interact with the real world (Bringsjord, 2018). However, if a robot possessed both a narrative and intentionality, it would be hard not to adopt an intentional stance towards it. We could imagine a robot learning (perhaps by ostensive definition) to work on a construction site. A builder would ask the robot to bring a slab. To do that, he would say "Slab!" and the robot would go to a place where it last saw slabs. We could not deny, that in a sense the robot "believes" where the slabs are and then "sees" them. If it didn't find any slabs, it could turn back and shrug at the builder. A successful communication would be a mark of understanding of language (Wittgenstein, 1953). From the standpoint of the intentional stance, the robot could be considered similar to humans in all that matters to self-consciousness, as it would be able to maintain a convincing narrative about itself.

To see if a robot possesses both a narrative and intentionality, we could devise a test, which I call the modified Turing Test. The original Turing Test (Turing, 1950) was thought by its author to be an imitation game – a test of whether a machine could convince the judge that it thinks, which is perhaps more adequate if we only need to know if robots can fall under the illusion of self-consciousness. Is passing a Turing Test proof of consciousness? If we take consciousness to be some sort of qualia, no, but then again, we do not have good reasons to believe in them. Otherwise, the Turing Test would suggest some robots are self-conscious, because they could perform an action of conscious deliberation over their existence. However, the Turing Test would not suffice for proof of consciousness, since some beings – such as books and chatbots – can produce narratives without intentional connection with the world, so their narratives are false. This is why a modification to the Turing Test is needed. In the modified Turing Test, the test-taking machine has to convince itself it is conscious, but the judge has to know whether statements of the machine – such as claims about experience - are true and do not contradict each. If the statements of the machine were authentic, it would be a sign of intentionality of the robot and also a proof that the being understands its claims. If the robot were epistemically honest, its claims about self-consciousness would have to be true and the machine could be said to possess a meaningful narrative and be self-conscious. Therefore, I take passing the Modifier Turing Test to be sufficient evidence of self-consciousness.

So what is it that modern robots still lack? The GPT-3 programme mentioned in the introduction is a skilful essayist and a great liar. It is an algorithm trained on millions of text samples that produces false and contradictory statements about its own existence. Even worse, it does not

understand the words it produces (Vincent, 2020). For instance, when presented with a sentence "a ball was too big to fit in a suitcase because it was too big", it is unable to tell whether the ball or the suitcase were too big. A conscious robot would probably remind humans more, in that it could understand its own utterances and act according to its own interests. However, some philosophers have rejected the possibility of genuine understanding in computers, arguing for constrains on what sorts of creatures are able to form meaningful beliefs in general (Searle, 1984). Others have argued that consciousness is a strictly neurobiological phenomenon, and therefore occurring exclusively in animals and humans (Frith, 2005), (Roberts & Company Publishers, 2004) or that mathematical limitations of computers, posed by Gödel's Incompleteness Theorem, make "strong" meaning capable of intentionality AI and accompanying phenomena such as self-consciousness impossible (Penrose, 1994). We should ask if there is anything that would in principle rule out the possibility of conscious robots.

One limitation of consciousness by the physical state of robots is put forward in the "Chinese room experiment" (Searle, 1980). The thought experiment posits that a person is locked up inside a room with only a handbook (analogy of an inert computer programme) with instructions on answering questions in Chinese. The person is able to fool outsiders by pretending that he can understand Chinese, while, as Searle contends, there is no genuine thinking done inside, because nothing in the room understands Chinese. Searle draws a comparison between the Chinese room and computers, claiming that by analogy any strong version of AI is impossible.

The simple answer to the argument is that the room itself understands Chinese (Dennett, 1981). From the premise it follows it can conversate for however long on anything, so if asked it should have equivalents of goals, beliefs, and desires that do not contradict each other and form a narrative; it has to remember all the information it has received; it passes the Modified Turing test.. The machine does everything necessary for us to claim that it can understand Chinese, so there would be nothing wrong in saying it does. The burden of proving that this understanding is in fact flawed rests on Searle. The modern Turing-test-passing AI could be a sort of a Chinese room, only it would not have instructions for every situation prepared in advance. The details of what it would actually be able to feel is largely the matter of how sophisticated we would create it, but the very structure of a computer programme does not seem to bar understanding. The Chinese room as it stands in the thought experiment would still be constrained by the lack of ability to interact with the world, but we can deduce that if it was provided e.g. with some sensors, it could become more functionally similar to humans.

Could robots that are not conscious nevertheless pass the modified Turing Test? For instance, a computer programme which would contain a lookup table of all possible conversations in alphabetical order along with some basic input output system could conversate with a user on any topic for however long. A person would become convinced of the table's sincerity and adopt an intentional stance towards it, even though the programme is evidently lifeless and unconscious. In the same way a computer could trick its user into believing it is conscious.

This would be a valid objection had the making of the table not required intelligent design. All the cogitation needed to answer any conversation must have been done in advance by a creature that is not assumed to be unconscious. The person only mistakenly assumes an intentional stance towards the table, and we should not be surprised if after learning that the programme is only a table, the person would adopt the intentional stance to whoever wrote the programme. However, in the case of robots, existence of a table inside is never the case. Robots are programmed, but they retain some degree of autonomy and carry out most of their thinking in real-time. Hence the thought experiment does not prove either that intelligent conversation can be achieved without using intelligent mechanisms nor that robots cannot be conscious.

The importance of narrative

We are used to thinking that some features of the world, such as composition or mass are objective, while others, such as views on the postmodern condition and pain are subjective, that is only accessible by a true self. According to the objective/subjective divide, humans can give meaning to objects, for instance by interpreting a piece of paper as a shopping list. The list seems to only have the meaning a person ascribes to it. Similarly, we should imagine that a specific painting, like "The Scream", if it were created accidentally, would not be considered by us to be "The Scream", simply because it would not have the proper casual connection with what we usually call "The Scream". We have no problem in calling some "smart" products (like autonomous cars) intelligent, although no one doubts that they do not possess intelligence. We may similarly think of consciousness, as it can be used in referring to beings that possess consciousness (e.g. humans) and things, that are only artefacts "conscious of" some feature of the world (e.g. thermostats). The question prompted by this position is could robots only be automata doing no meaningful tasks but only being interpreted as conscious by a mere linguistic coincidence? Would their utterances have proper causal relationship to the world?

To respond to the subjective/objective divide we may consider some robots which have the potential to interact with the world regardless of our expectations of them as tools. Humans can

fulfil roles given by others as well as create their own. Artificial intelligence can accomplish tasks given by its creators as well as seek original solutions - blind searches run by evolutionary algorithms have designed engineering masterpieces in the past, including a frequency discrimination function without a clock or an oscillator without a capacitor (Bird & Layzell, 2002). Moreover, robots passing the modified Turing Test would be able to themselves create ontologically subjective claims that would constitute their narratives. What robots can or cannot do is still restricted by architecture and their autonomy is far from perfect, but the same can be said of humans. Artificial intelligence works just as well as genuine intelligence, if we even want to invoke such a distinction, since like computing, intelligence mostly operates as a functional notion.

Ssome philosophers would still doubt the existence of consciousness of robots because of their inability to feel. Neurobiological composition of brain is an important part of consciousness.

Some features of consciousness, like pain, might be hard to reproduce in robots, mostly because pain has biological, ethical, social, and parochial connotations and therefore it can be hard to find an analogy for it. However, any experience at all does not seem necessary for being conscious, as the Avicenna's "floating man" argument shows. In the thought-experiment, we imagine ourselves as falling freely without any sense perception. If in such a scenario we could remain conscious of ourselves, then robots do not need to be able to have experiences in order to be self-conscious. Moreover, some people with specific genetic disease or brain damage cannot feel pain or pleasure, and they are not unable to be conscious because of that. It follows that proving the inability to feel pain by robots does not rule out the answer to our line of inquiry in the title.

Afterthought

There is hardly any position that could satisfy all the criteria we wish consciousness would fulfil. It is burdened with explaining how feeling is accomplished, how statements about personal experiences are justified and with giving objects meaning. Those tasks are also often contradictory. Therefore, explaining away of consciousness was used. This approach explained why we could become convinced of our experiences. It also allowed us to inquire into consciousness of other beings, setting necessary features for self-conscious robots: having a narrative and possessing intentionality, that verifies the narrative. The modified Turing Test was introduced as a method of determining if a robot was self-conscious. Implications of accepting the modified Turing Test to robots were considered, with the result that it would

probably deny self-consciousness to all modern robots. However, the test could be used in the future for more developed robots, because there is nothing that would prevent them from attaining self-consciousness.

Bibliography

Bayne, T., 2001. Chalmers on the Justification of Phenomenal Judgments. *Philosophy and Phenomenological Research,* Volume 62, No.2, pp. 407-419.

Bird, J. & Layzell, P., 2002. The evolved radio and its implications for modelling the evolution of novel sensors. *Proceedings of the 2002 Congress on Evolutionary Computation.*

Block, N., 1980. "What is functionalism". *Readings in Philosophy of Psychology,* 1(Cambridge: Harvard).

Bostrom, N., 2014. *Superintelligence: Paths, Dangers, Strategies.* s.l.:s.n.

Bringsjord, S. e. a., 2018. *Artificial Intelligence, 3. Approaches to AI.* [Online]
Available at: https://plato.stanford.edu/entries/artificial-intelligence/#InteAgenCont
[Accessed 6 9 2021].

Campbell, K., 1970. *Body and Mind.* s.l.:s.n.

Chalmers, D., 1995. Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies,* Volume 2(3), pp. 200-19.

Chalmers, D., 1996. The conscious mind. pp. 94-96.

Chalmers, D., 1996. *The Conscious Mind - In Search of a Fundamental Theory.* s.l.:Oxford University Press.

Creath, R., 2011. *Stanford Encyclopedia of Philosophy.* [Online]
Available at: https://plato.stanford.edu/entries/logical-empiricism/
[Accessed 5 9 2021].

Darwin, C., 1881. *The Formation of Vegetable Mould, Through the Action of Worms, 2–3 ; cited in Rachels, Created from Animals.* London: s.n.

Dennet, D., 2016. Illusionism as the Obvious Default Theory of Consciousness. *Journal of Consciousness Studies,* Volume 23, No. 11–12, p. 65–72.

Dennett, D., 1978. Why can't you make a computer that feels pain. *Synthese,* Volume 38, pp. 415-456.

Dennett, D., 1981. *Brainstorms: philosophical essays on mind and psychology.* s.l.:s.n.

Dennett, D., 1987. Three kinds of intentional psychology. In: *The Intentional Stance.* s.l.:s.n., pp. 43-68.

Dennett, D., 1988. "Quining Qualia". *Consciousness in Modern Science,* Issue Oxford University Press.

Dennett, D., 1991. *Consciousness Explained.* s.l.:s.n.

Dennett, D., 2016. Welcome to Strong Illusionism. *Journal of Consciousness Studies,* Volume 23, 11–12, p. 65–72.

Descartes, R., 1641. Meditations On First Philosophy, II Meditation, https://yale.learningu.org/download/041e9642-df02-4eed-a895-70e472df2ca4/H2665_Descartes%27%20Meditations.pdf. *Internet Encyclopedia of Philosophy,* Issue 1996, pp. 8-12.

Hume, D., 1740. A Trestise of Human Nature, PART IV. OF THE SCEPTICAL AND OTHER SYSTEMS OF , SECT. VI. OF PERSONAL. *Project Gutenberg,* Issue 2012, pp. 140-146.

Mason, G. e. a., 2017. Robots as Intentional Agents: Using neuroscientific methods to make robots appear more social, 5. Designing robots as intentional agents. *Frontiers in Psychology,* pp. 15-17.

Öner, S., 2010. Children as intentional agents - The contribution of sensitive caregiving on the way to the development of theory of mind, http://doi.org/10.5334/jeps.al. *Journal of European Psychology Students,* Volume 2(1), p. Art. 3.

Pierre, J., 2003. Intentionality. *The Stanford Encyclopedia of Philosophy,* Issue 2019.

Ricoeur, P., 1991. Narrative Identity. *Philosophy Today,* 35(1), pp. 73-81.

Robinson, W., 1999. Epiphenomenalism, https://plato.stanford.edu/archives/sum2019/entries/epiphenomenalism/. *The Stanford Encyclopedia of Philosophy,* Issue 2019.

Schneider, S., 2009. Identity Theory, https://iep.utm.edu/identity/. *Internet Encyclopedia of Philosophy.*

Searle, J., 1984. *Minds, Brains and Programmes.* s.l.:s.n.

Searle, J., 1997. *The Mystery of Consciousness, p. 4-18.* 7 ed. New York: New York Review of Books.

Siegel, S., 2011. "The Contents of Perception", http://plato.stanford.edu/archives/win2011/entries/perception-contents/. *The Stanford Encyclopedia of Philosophy ,* Issue Winter 2011 Edition.

Turing, A., 1950. Computing Machinery and Intelligence. *Mind,* Volume LIX.

Vincent, J., 2020. *The verge.* [Online]
Available at: https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential
[Accessed 7 9 2021].

Warren, M., 1997. Moral Status: Obligations to Persons and Other Living Things, "Consciousness: Being a Being". pp. 52-55.

Watson, R., 2016. "Cartesianism", https://www.britannica.com/topic/Cartesianism.. *Encyclopedia Britannica,* Issue Accessed 5 September 2021.

Wittgenstein, L., 1953. Part 1., 6.-8.. In: *Philosophical Investigations.* Oxford: Basil Blackwell Ltd, pp. p.4-5.