Presented at the DLSU Research Congress 2019
De La Salle University, Manila, Philippines
June 19 to 21, 2019

Programming the Consciousness of a Non-living Thing: A Freudian Analysis of Artificial Intelligence

Kim Nucom
Philosophy Department
De La Salle University
kim_nucom@dlsu.edu.ph

Abstract: A general definition of a robot is a machine programmed to carry out certain tasks or actions. While industrial robots have their place in society, the popularity of artificial intelligence creates a narrative around the concept of humanoid robots and the limits of technology when it comes to creating or transferring consciousness. The question is how far we can go with creating a robot that is as human-like as possible. For a robot to have a mind akin to that of a human's, it must be capable of recognizing moral values and generating its own thought process the way a typical human does. In this paper, I shall be using Freud's theory of the tripartite psyche to examine how humanity would fare in a robot whose traits and actions are entirely programmed. My objective is to see if we can program consciousness such that a robot can fully act on its own volition without direct human intervention. With that in mind, I also enquire into the ethics of robot consciousness and its consequences on society if such a robot came into existence. Many questions on morality such as the treatment of robots with human minds would have to be resolved, which our society might not be prepared to answer. I hence look into the repercussions of artificially creating a psyche, especially if it's modeled on the human id, ego, and superego.

Key Words: consciousness; Freud; human; psyche; robot

1. INTRODUCTION

Perhaps the origin of the concept of conscious robots came from the creation of the Turing machine almost a century ago. A pioneer in challenging the mystery of human thought, Alan Turing's legacy paved the way to the development of computers and the study of artificial intelligence (Takeno, 2013). Today, robots have an important place in the field of technology and engineering. More often than not, they are used in industry to do hard labor that normal humans aren't capable of doing. Typically they're employed in factories to do repetitive tasks for maximum speed and efficiency. Sometimes they're also used as a cheaper alternative; one robot could be capable of doing the work of fifty humans at once at a much lower cost.

A robot's degree of autonomy is measured by its ability to perform the ordered actions (Xie, 2003). When it comes to the topic of artificial intelligence, this notion is very popular, particularly in the realm of science fiction. The concept of humanoid robots is currently being explored by the technology sector, with most notable examples being Erica, who anchored a news segment in Japan (Dar, 2018), and Sophia, the world's first robot citizen (Stone, 2017). Now these humanoid robots have human-like bodies. However, do they also have human-like minds? In this paper, I try to answer this question using both a philosophical and psychological approach. I chose to employ Sigmund Freud's famous theory of the three-part human psyche because it easily and directly relates to the following questions: (1) Do robots have feelings, impulses, and emotions? (2) Can they generate a thought process that is similar to that of a typical human's? (3) Is there a concept of morality imbued in a robot, a purely programmed and artificially created machine?

2. THE TRIPARTITE PSYCHE

In order to understand the robot psyche, we use the human psyche as the model for comparison. Following Freud's theory of the psyche, it consists of three major systems: id, ego, and superego. These three work hand in hand to fulfill a human's basic needs and desires (Hall, 1954). When the psyche is healthy, the human creates positive interactions with their environment as compared to one with an unhealthy mind, whose dissatisfaction affects the way they approach the world. This, we witness in our everyday lives, in our society. On the other hand though, what does this entail for robots? Assuming they were programmed to behave in a human-like manner, one could not help but ask if robots can actually fully reach the level of human mental faculties. I shall put forth then a discussion on whether Freud's theory about humans can apply just as well to robots.

2.1 The Id

The strongest part of the psyche according to Freud, the id is the seat of our instincts, the part that contains our everyday drives and impulses. It is guided by the Pleasure Principle (Mansukhani, 2013b), in which its primary aim is to experience pleasure as much as possible and minimize pain just as much. It wants to rid the person of tension—tension being defined as pain or discomfort which can only be relieved by pleasure or satisfaction (Hall, 1954). The id will do anything to get rid of this feeling of displeasure, following its own instincts and ignoring the laws. Among the three systems of the psyche, it is the most in tune with the body and most away from the mind.

The id cannot be changed with experience because it acts without influence from its environment or external world (Hall, 1954). With this in mind, it already fails the first criteria of the human-like robot. For robots can still be altered and reprogrammed and thus, its instincts cannot be considered innate. This also begs the question of robots having the concept of instincts. In humans, we have hunger, thirst, lust, passion, and impulse. Firstly, a robot cannot naturally crave or thirst in both the mental and physical sense, for it is not structured in the same way as a human. The robot is a mere machine; if it can be controlled as we control machines, that makes it seem like a machine itself (van de Vate, 1971). A robot can have impulses and its own idea of pleasure, but those in themselves are programmed into the robot. The limit of this kind of programming is that a robot can only learn so much from its external world from how much it was programmed to learn. Because the id and its essence is not naturally inborn in a robot, said robot will reach a point wherein it will no longer take any new impulses. To put it simply, a robot can never have the same drive as a human.

2.2 The Ego

The ego is the voice of reason, acting as a mediator between the id and the superego. It is, Freud says, guided by the Reality Principle (Mansukhani, 2013b). The aim of the reality principle is to postpone an action until the actual object that will satisfy the need

has been found (Hall, 1954). This postponement or action, Hall adds, means that the ego has to have the ability to tolerate tension or feelings of displeasure until it can get rid of them in an appropriate manner and place. Unlike the id, the ego takes note of its environment before making decisions. The role of the ego is control the id's impulses as much as it can, and stop it from going wild or doing something inappropriate. In all aspects, the ego is born out of the id (Hall, 1954).

Now let's assume the existence of the id in a robot. Suppose that robots indeed had impulses and passions. How does the ego counter this? By of course employing its power of governance and pushing down the id's strong instincts in order to align it properly with the environment. But then a robot programmed to do this simply cannot showcase any impulses. For if it is designed such that it has the ego, it will always perform perfectly and in line with its ordered tasks. In short, there's no id to control. But say there's an imaginary scenario wherein a robot's id and ego are measured in random or externally generated percentages, and whoever has the higher correlates to the robot's next action. As much a good idea this is, we just cannot accept this as a perfect model of the ideal human psyche for it simply doesn't work that way. There are many instances in our lives when our ego won against the id, even if the id is more overwhelming of the two at a given moment. This is the Reality Principle doing its work. A robot intending to copy a human also cannot just choose between two alternatives at random without a sense of reason governing it.

2.3 The Superego

The superego is, as we put it, the moral branch of the psyche. It enforces rules and laws in order to prevent us from acting impulsively. Whilst the id looks for pleasure and the ego looks at reality, the superego strives for perfection; it represents the ideal rather than the real (Hal, 1954, p. 31). It's born out of social interactions with other people, from the moment we get exposed to the notions of good and evil as a child. It's the reason why we feel guilt or shame whenever we fall for the id's impulses and experience the repercussions on the ego. Now Freud believed that the superego has two parts. The part that makes us feel guilty is what we call the conscience. The other part is the ego-ideal, which refers to our internalized standard of perfection.

It is with this that I argue that a robot cannot have a standard of perfection beyond its programmed design, for a robot is not capable of distinguishing between right and wrong in the human sense. It can learn to distinguish which action to take based on the input it takes from its environment, but it can't decide for itself if it's reached a certain level of goodness or perfection.

According to Bittle (1953), one main source of knowledge is experience. Experience has two parts. We have consciousness, which is our awareness of our mental states, and sense-perception, which enables us to apprehend the world around us. Consciousness is internal experience, while sense-perception is external experience. Now a properly programmed robot is capable of sense-perception. It can make decisions based on its environment. On the other hand, consciousness is not something that is readily available to the robot. It can process things internally in the programming sense, but it cannot evaluate things as good or bad or just or unjust or any other moral adjectives. It cannot formulate its own sense of morality.

3. ROBOT CONSCIOUSNESS

The word 'conscious' comes from a Latin word that means "knowing things together" (Marsh, 1977). In this section we question if a robot can truly have the epitome of a human consciousness, given that it's failed the criteria for Freud's human psyche. It's important to take note that the original purpose of humanoid robots is to test their intelligence—intelligence, in this sense, meaning the ability to link perception to actions for the purpose of achieving an intended outcome (Xie, 2003). For now, we're testing the limits of artificial intelligence by trying to program the perfect humanoid robot. However the problem with this is that programming has a problem-solving nature that isn't sufficient for the objectives at hand. At the moment, we cannot program concepts such as feelings and emotions because they are not tasks or problems to be solved by the robot (Gunderson, 1968).

To create a fully conscious robot with complete human faculties is but a huge feat we have yet to do. There are many ways to achieve this; one way is by understanding how languages work. According to formal logic, human thought arises from languages (Takeno, 2013). Studying artificial intelligence also entails studying languages. Chomsky thought that creating a machine that understands and speaks natural languages was an important means to understanding the human mind. (Takeno, 2013). But the main issue at hand

is regarding our full knowledge of the mind, or the lack thereof. Most philosophers who speak about the nature of the human psyche fail to confront the fact that a large portion of the mind is not accessible to us (Mansukhani, 2013a). Majority of our mind is part of the unconscious world, filled with our deepest desires and drives that we may or may not be aware of until we delve through the caverns of our consciousness. So one could argue if a robot can truly have the concept of an unconscious. In the scenario that it can, we question how such could be created. Do we program the unconscious or does the unconscious develop out of the robot's mental faculties? This is a hard question to answer for even we cannot provide a solution to this problem regarding the human psyche.

But assuming we did find out that robots can have a consciousness, what does that entail for these said robots? If they are granted with human-like minds, that means we have to start treating them the way we treat fellow humans. This raises the question regarding the rights of humanoid robots with human intelligence, if, say, they should be made citizens like Sophia the robot. It is also safe to assume that intelligent human-like robots can develop their own moral compass, accompanied with impulses, drives, and emotions. If that is so, this makes them subject to the human laws, for it signifies that they can recognize different notions of good and evil, meaning they cannot be excused from social norms and constructs. But humans then also are obliged to treat them as fellow humans, regardless of their bodily structure.

But what are the consequences though of robots becoming humans in the metaphysical sense? Questions arise when we think about giving them the liberty to roam free in our society. If normal humans are capable of destruction, imagine the power of robots who have mechanical abilities far more than that of a human's. Back in October 2007, a semi-autonomous robot cannon deployed by the South African army malfunctioned, killing 9 soldiers and wounding 14 others (Wallach & Allen, 2009). Imagine the potential scenarios if such machines became fully autonomous. Now whether we like it or not, autonomous robots are an incoming bundle in our future. Would they adhere to the tripartite psyche proposed by Freud? That, we cannot be certain as of yet. But it is in our hands to make sure these robots don't cause any harm to society, just as much as we govern our fellow humans. Engineers are foremost held responsible for the creation of these robots. If giving machines moral standards would improve public welfare and safety, then engineers are obligated to make it happen (Wallach and Allen, 2009). We are liable for the robots we manufacture, for the consciousness we create.

4. CONCLUSION

To conclude, I have shown that the robot psyche is but a myth using Freud's theory of the three-part personality. However, I cannot deny there are feasibly better models out there that could argue better for the existence of the robot psyche. In the chance that a robot consciousness could indeed exist, we must be prepared for the changes that this phenomenon would bring to our society.

5. REFERENCES

Bittle, C. N. (1953). *Reality and the mind: Epistemology*. Milwaukee, WI: The Bruce Publishing Company.

Dar, P. (2018, February 12). A robot called Erica set to become news anchor in Japan. Retrieved from https://www.analyticsvidhya.com/blog/2018/02/erica-robot-news-anchor-japan/

Gunderson, K. (1968). Robots, consciousness, and programmed behavior. *The British Journal for the Philosophy of Science*, *19*(2), 109-122.

Hall, C. S. (1954). *A primer of Freudian psychology*. New York, NY: Penguin Books USA.

Mansukhani, R. (2013). Delving into the depths of the psyche: An introduction to the philosophy of the unconscious. In L. dlR. Garcia (Ed.), *Exploring the philosophical terrain* (pp. 221-231). Quezon, Philippines: C & E Publishing.

Mansukhani, R. (2013). The self in conflict: Freud's theory of human nature. In L. dlR. Garcia (Ed.), *Tao po? Readings in philosophy of person* (pp. 36-52). Quezon, Philippines: C & E Publishing.

Marsh, C. (1977). A framework for describing subjective states of consciousness. In N. E. Zinberg (Ed.), *Alternate states of consciousness* (pp. 121-144). New York, NY: The Free Press.

Stone, Z. (2017, November 7). Everything you need to know about Sophia, the world's first robot citizen. Retrieved from https://www.forbes.com/sites/zarastone/2017/11/07/

everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/#7494bfe046fa

Takeno, J. (2013). *Creation of a conscious robot: Mirror image cognition and self-awareness*. Singapore: Pan Stanford Publishing.

Van de Vate, D., Jr. (1971). The problem of robot consciousness. *Philosophy and Phenomenological Research*, *32*(2), 149-165. doi:10.2307/2105945

Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. New York, NY: Oxford University Press.

Xie, M. (2003). *Fundamentals of robotics: Linking perception to action*. Singapore: World Scientific Publishing.