"Could you be friends with a robot?"

With the advancements in technology and artificial intelligence, it seems likely that in the future there will be countless interactions between humans and robots. Over time, client-facing workers such as postmen, cashiers, butlers and so on, will be replaced by robots and automated machines. These interactions will inevitably lead to a relationship with these robots. This begs the question, is it possible to have a relationship with a robot? If yes, then, is it possible to be friends with a robot? I assert the view that it is not possible to be friends with a robot. While philosophers such as Dennett and Minsky have influenced my thoughts significantly, it has become evident through other thinkers such as Searle and Chalmers that it is not possible for robots to have a consciousness. But, before we delve any further, it is paramount that we establish some definitions. What is friendship? And, what defines a robot?

Google defines friendship as "A relationship of mutual affection between people." Although this is a satisfactory description, it is rather vague. It is important to find a deeper and more meaningful definition to truly understand what consists of a genuine friendship. I believe that the term "friendship" is best illustrated by the ancient Greeks due to their ability to be much more expressive and precise in their discussions about relationships. For example, they had several words for 'love' (some that we will look at in this essay), like "agape", "eros" and "philia". In contrast, the English language only has the one general term and lacks specificity and nuance. 'Agape' love is unconditional, undeserved love, the type that God has towards even his most sinful creation. 'Eros' love has sexual connotations and was more commonly used to describe intimate or erotic love (the word 'erotic' actually being derived from 'eros'). Finally, 'philia' is considered a friendly feeling of compassion towards someone else; it is more affectionate than intimate. In the remainder of the essay, as we are concerned about whether robots can be friends with humans, we will be interested in philia as this is the closest and most precise definition of friendship. Thus, is it possible to have philia relationships with robots?

First and foremost, how can we identify philia relationships? In Aristotle's 'Nicomachean Ethics' (book VIII), the great Greek Philosopher highlights the three subcategories of philia; friendship for utility, friendship for pleasure and friendship for virtue. A utility friendship is a relationship between individuals whose foundation is gaining some value, physical or non-physical, out of the other party members. For example, being closer to someone because of their wealth or status. A pleasure friendship is a relationship between individuals whose foundation is the pleasure derived from the interactions between the members involved. Examples include someone you enjoy participating in a shared activity with, such as a sport or a hobby. Finally, a virtue friendship is a relationship between individuals whose primary merit lies in the enriching effect it has on the virtues of those involved. Aristotle observes that pleasure and utility friendships are, at most, deficient forms of friendship due to their limitations, both temporal and potential. Temporally, friendships based on utility or pleasure will end when their objects end. A business opportunity will end if you find a better one. A friendship will end if you meet someone more interesting and exciting. Furthermore, utility and pleasure friendships may blind us to the greater potentiality and worth of a person. Virtue friendships complete the intended telos (purpose) of human relationships. Temporally, they are not limited to maintaining pleasure or utility but are suited to last a lifetime and, therefore, one should adopt the view that the only genuine friendships are ones based on virtue.

Contemporary British philosopher, Julia Annas (1988, p. 1), summarises Aristotle's virtue friendship. She says:

> A friend, then, is one who wishes and does good (or apparently good) things to a friend, for the friend's sake, wishes the friend to exist and live, for his own sake, spends time with his friend, makes the same choices as his friend and finds the same things pleasant and painful as his friend.

The key point highlighted is the importance of having mutual interests and shared values. Selflessness is also a principal theme as there should be no ulterior motives, but purely participating in the relationship “for the friend’s sake”.

Now that there is some clarity as to what principles and grounds I will be assessing the question from, it will help to display the essay question in a logical format. A simple way of laying out the structure of the essay is to identify our premises and our conclusion (A premise is what you are arguing from; it is a fact or piece of evidence that you offer to establish your conclusion. The letter “P” will be used to denote a premise and “C” for a conclusion.)

P1 Aristotelian virtue friendships (the only true type of friendship) require mutual, shared interests; genuineness; and equality.

P2 Robots can never fulfil these requirements.

C1 Therefore, we can never have Aristotelian virtue friendships with robots.

This is a deductive argument - an argument where if the premises are true, the conclusion must necessarily be true. Now say, hypothetically, that P1 and P2 were true, then it would consequently follow that the conclusion must also have to be true. Therefore, from a validity perspective, this argument is as valid as it gets as there are no flaws in the logical flow of the premises and conclusion. But it is the soundness of this argument where the debate lies. An argument is sound if the premises are genuinely true in our world, not hypothetically, but metaphysically (in reality). P1 is true in our world as it is an analytic statement as, a relationship that “requires mutual, shared interests; genuineness; and equality” is just the definition of a virtue friendship. All that we have left to decide is whether P2 is metaphysically true. Once we have come to a conclusion, then we can know whether this is a sound argument or not. So the question has now become, can robots have relationships that are mutual, genuine and equal?

To answer this question, we must focus our attention on what it means to be an artificially intelligent being. According to Peter Norvig and Stuart J. Russel (2009, vii), two computer scientists and joint authors of “Artificial Intelligence: A Modern Approach”, AI is “the study of agents that receive precepts from the environment and perform actions.” These robots are “intelligent agents” that gather information from their surroundings and act accordingly. Some philosophers would argue that to be able to do all this, robots must have an inner conscience; they must be alive. Consciousness is the ability to understand and relate with other beings as well as having an inner mental life.

I believe that robots cannot have a virtue relationship with humans because they do not have an inner mental life. Manufacturers will make it seem as though these robots have human emotions and empathy, but these are just marketing ploys used by companies to sell more of these robots. They cannot truly communicate with us, they cannot truly experience with us and they do not have their own values and feelings. Consciousness is needed to portray these qualities. Consciousness gives one the ability to relate to other beings and to truly be alive. This reasoning is justified by David Chalmers (2010, p. 106-109). Chalmers is a dualist; he believes that the mind and body are separate. He believes that the mind, or consciousness, is like an “inner movie”, a first-person film. Chalmers gives a very interesting argument against physicalism (the belief that consciousness and everything in existence can be explained using only physical interactions such as neurons and synapses). He developed the idea of a philosophical zombie; a being that is identical to us in every physical aspect and will give the same response to a given stimulus as us humans but lacks consciousness. He also stated that for phenomenal concepts (such as talk about consciousness), conceivability implies possibility. His argument is as follows: physicalists believe that every single thing can be explained

through material, physical properties. They believe that even consciousness can be explained by neural connections and chemicals. Therefore, if we make an exact physical duplicate of our world, according to a physicalist, everything in this duplicate world should be identical to our world (including the ability to have conscious beings). Chalmers' response is that he can conceive of a world that is a physical duplicate of ours, where the beings inside it lack consciousness and are philosophical zombies. Therefore, according to Chalmers, as this world is logically possible, it must be metaphysically possible. Thus, consciousness cannot be explained by just physical properties and clearly, something immaterial and non-physical is required to illustrate this phenomenon. This immaterial, non-tangible, temporal requirement is the mind and so physicalism is false.

An outline of his argument would be:

P1 According to physicalism, everything that exists in our world (including consciousness) can be reduced to physics interactions.

P2 Thus, if physicalism is true, a metaphysically possible world in which all physical properties are an exact duplicate of our world must contain everything that exists in this world. Therefore, conscious experience must also exist in this duplication.

P3 In fact, we can conceive of a world physically indistinguishable from our world but in which there is no consciousness (a zombie world). Therefore it must mean that such a world is metaphysically possible.

C1 Therefore, physicalism is false.

If Chalmers is correct in his view, even though it is unlikely that in the next hundred or two hundred years we could create such advanced robots that are an exact replica of our own physical properties, it would not matter as they still won't have an inner mental life. And so without consciousness, these robots simply cannot fulfil the requirements of a virtue friendship.

However, Chalmers' argument does not fully convince me. Physicalists such as Dan Dennett (1995, p. 322-326) argue that while consciousness and subjective experience exist in some sense, they are not as the proponent of the zombie argument claims. Dennett believes that consciousness is an extremely complex phenomenon and that other thinkers have greatly underestimated its intricacy. He asserts that taking away the ability to experience pain, for example, cannot occur without there being any behavioural changes as a consequent. So in this duplicate world, if these zombies are exact physical replicas of us and they give the same responses to stimuli as us, then it must be that they also have a consciousness. Consciousness and behavioural responses go hand in hand for Dennett. Therefore, the very idea of a philosophical zombie is contradictory. Consequently, these zombies will be capable of having virtue friendships with humans and Chalmers' argument falls apart. This is a very strong response to Chalmers' zombie argument as Dennett's idea that consciousness is a truly complex phenomenon is very convincing. If, in the future, we are capable of manufacturing robots that are an exact physical duplicate of ourselves (although, a very unlikely prospect), then these beings will have consciousness and so can have virtue relationships with humans.

Another response to Chalmers that convinces me towards the view that you can be friends with a robot, is an argument from Marvin Minksy (1998, p. 2). Minsky proclaims that Chalmers' argument is circular in nature. The proposition of the possibility of a being without consciousness and one that is physically identical to us assumes that it is not physical interactions that cause this subjective experience. But, this is exactly what the argument is trying to prove, that subjective experience is not the result of physical interactions. This highlights a key problem with Chalmers' thinking; he uses his concluding ideas at the beginning of his argument in his very definition of a philosophical zombie. This is a logical fallacy and illustrates the weakness of his reasoning. Thus this is very convincing towards

the belief that consciousness can be explained by physical elements and therefore, using these physical elements, we can construct a robot with consciousness allowing it to have virtue friendships with humans.

However, in this dualism versus physicalism debate, the most persuasive argument is made by a philosopher called John Searle (1984, 'Minds, Brains and Science'). Searle asks us to imagine a person who only speaks English, who is locked in a room. Inside he has a rulebook that tells him that if he receives a certain Chinese symbol through the door, he should return another Chinese symbol out of the ones he already has in the room with him. The person sending the Chinese symbol in the room is fluent in Chinese. To this person outside, she is having a fluent conversation in Chinese as the man inside is providing the correct outputs based on the inputs he gets according to a rulebook. But Searle has shown that even though we see the correct inputs and outputs, we know the man has no real understanding as he does not know Chinese. This is exactly how a computer works; it is not thinking and it does not understand. It merely has programs running that give the right outputs to certain inputs but it does not have mental states. Consciousness and the mind are not like this; they do not work like a computer. The mind understands. Systems which function based on algorithms cannot comprehend any meaning or operate with intentionality. Therefore, it is clear to me that an Artificially Intelligent robot, no matter how advanced, can never achieve consciousness. Thus, we cannot be friends with a robot.

Other philosophers may present a different perspective. They argue that although, in the present, it seems unlikely that robots that have inner mental lives can exist, there is a high possibility that in the future they could. We will soon be able to manufacture robots with very sophisticated mental structures to go beyond mimicking our behaviour to actually acting more human-like. The idea that in the future we will be able to construct a machine with a conscience and a sense of authenticity does not seem absurd at all. Although this argument makes logical sense, it does not convince me, as we do not have knowledge of the future and so it is uncertain whether this could ever be possible. Thus, this argument seems to be avoiding the question and it does not give an applicable or tangible case that we can properly scrutinise.

Alternatively, one could take a hedonistic approach to show that robots can have a virtue relationship with humans. According to hedonism, the pursuit of happiness is the most important thing, it is the highest good and the ultimate telos for humans. Therefore, it does not really matter whether the robot actually has an inner mental life, as long as it behaves in such a way that it seems like it is satisfying the conditions of a virtue relationship. This is because, at the end of the day, this will give us the same happiness as if the robot actually did have these qualities. This is further reinforced by the fact that virtue friendships with humans are no different. You may have someone in your life who pretends to share your interests and values but is actually lying to you. We cannot tell if this person is lying, so from our perspective, we would still say we are in a virtue relationship with them as we are none the wiser. Therefore, if you get the same pleasure from a robot merely acting like it has an inner mental life and from a robot that actually does, then is there really any difference? They are both, in your eyes, a virtue relationship. Thus, it seems that you can be friends with a robot.

Yet, I still remain in favour of humans not being able to be friends with a robot. Robert Nozick (1974, p. 43) has an interesting response to hedonism that is much more convincing. His "Experience Machine" argument says to imagine that you are in a simulation that is indistinguishable from reality. Whatever pleasure you feel in the machine is the same pleasure you would feel if you did the same action in this world. According to hedonism, these two actions are identical as they both yield the same amount of happiness. But Nozick points out that this is absurd. Surely, it's better to actually experience the action in the real world rather than having the mere illusion of experiencing it. Therefore, he concludes, hedonism is false. This is a strong argument as it seems almost ludicrous

saying that doing an act in a simulation is indistinguishable in value as doing the same act in real life. Consequently, when you apply this argument to the question in debate, it does matter whether robots actually have an inner mental life as virtue relationships need to be genuine and not a mere illusion. So since robots do not have an inner consciousness, they cannot provide these qualities, so you cannot have a virtue relationship with them.

To conclude, using Aristotle's three types of friendship, I illustrated that only virtue friendships are true and genuine. Then I formed the essay question into a deductive argument where we had to prove whether premise 2 (that robots can have mutual, shared interests with us) was a metaphysical truth in our world. To decide whether this was the case, I defined what it means to be a robot and whether robots can have inner mental lives as consciousness is required to fulfil the prerequisites of a virtue friendship. Some philosophers argue that robots may develop this inner conscience in the future when we have access to more advanced technologies. However, this is a weak argument as it is avoiding the question and almost giving up in providing an actual answer. Alternatively, Hedonists proclaim that robots do not even need to have inner mental lives, as long as they portray the values of a virtue friendship from our perspective. However, this was countered by Nozick's experience machine and how it is absurd that one would choose synthetic pleasure from a mere illusion over real-life pleasure. Dualism and physicalism also played a major part in the essay. If dualism is true, then the mind is separate from the physical and so a robot can never have consciousness. On the other hand, if physicalism is true then consciousness can be explained by physical interactions, so robots can have inner mental lives. Chalmers' zombie argument against physicalism seemed very weak to me as Dennett's counter, that consciousness is much more complicated than Chalmers' illustrated, is very convincing. Furthermore, Minsky's claim that Chalmers' zombie argument was cyclical in nature highlighted the flaw in the structure of his reasoning. However, Searle's Chinese Room influenced my conclusion the most. His concept of consciousness being the ability to understand, and his belief that a syntactic system that can only assess inputs and outputs (i.e. a robot) can never understand, seemed to make logical sense. Thus, I have come to a closure surrounding my stance on this question. I strongly believe that premise two is true and premise one was already proven true, meaning that the conclusion must necessarily be true. Humans cannot be friends with a robot.

REFERENCES

- Andrew Moore, (2013). Hedonism [online]. *Stanford Encyclopedia of Philosophy*. [Viewed 07 May]. Available from: https://plato.stanford.edu/entries/hedonism/
- Bennet Helm, (2017). Friendship [online]. *Stanford Encyclopedia of Philosophy*. [Viewed 01 May 2020]. Available from: https://plato.stanford.edu/entries/friendship/
- John Danaher (2017). Aristotelian Friendships and Robotics [online]. *Institute of Ethics and Emerging Technologies*. [Viewed 05 May 2020]. Available from: https://ieet.org/index.php/IEET2/more/Danaher20170225
- Philosophy Vibe, (2018). Functionalism and John Searle's Chinese Room [YouTube]. *Philosophy Vibe.* [Viewed 18 May 2020]. Available from: https://www.youtube.com/watch?v=EW5K1CyegJ0
- Robert Kirk, (2019). Zombies [online]. *Stanford Encyclopedia of Philosophy.* [Viewed 15 May 2020]. Available from: https://plato.stanford.edu/entries/zombies/

- Selmer Bringsjord and Naveen Govindarajulu, (2018). Artificial Intelligence [online]. *Stanford Encyclopedia of Philosophy*. [Viewed 03 May 2020]. Available from: https://plato.stanford.edu/entries/artificial-intelligence/

Zain Mobarik