

Kobe C. Lee

October 23, 2018

### **Machines Do Not Understand Happiness**

In the book *Anarchy, State, and Utopia* published by philosopher Robert Nozick in 1974, Nozick proposes a machine (the Experience Machine) that could theoretically provide any human connected to it with experiences that would make them feel unending happiness. This machine poses an issue for utilitarianism in theory, since, when we see pleasure or happiness as the overall good, and yet most people would decide not to plug into this machine, then there must be a greater pleasure than our hedonistic instincts. Robert Nozick is wrong. The experience machine does not provide an issue for utilitarianism, however not because it does not properly challenge happiness or hedonism, rather, it does not challenge utilitarianism due to the fact that the definition of happiness as noted by many utilitarian's is incorrect, as well as the fact that for any form of machine to understand human happiness, it must transcend the boundary between man and machine.

Now, Nozick gives his argument along a very specific set of lines, those being:

- Premise 1: If experiencing as much pleasure as we can is all that matters to us, then if we will experience more pleasure by doing x than by doing y, we have no reason not to do x rather than y.
- Premise 2: We will experience more pleasure if we plug into the experience machine than if we do not plug into the experience machine.
- Conclusion 1: If all that matters to us is that we experience as much pleasure as we can then we have no reason not to plug into the experience machine. (P1&P2)
- Point 3: We have reason not to plug into the experience machine.

- Conclusion 2: Experiencing as much pleasure as we can is not all that matters to us (*Modus tollens* used in Counter 1 and Point 3)<sup>12</sup>

Using this Criteria, Nozick comes to the correct conclusion that we, as human beings, would traditionally not plug into the experience machine. Where Nozick is wrong in thinking that this proves that there are things in life we value more than happiness. Nozick is mistaken in the idea that his machine would provide true happiness, outside of physical and chemical stimulants to *simulate* happiness. The machine that Nozick provides does not pose an issue for utilitarianism, when the proper definition of happiness is used in this experiment, or a non-hedonistic definition is used. However, for forms of utilitarianism that define pleasure as utility, this thought experiment completely invalidates their argument, if this is performed in a vacuum with the subject having no preconceived bias's or notions. As that is not the case, even hedonist-based utilitarian's may counter by putting forth the idea that any person, without any status-quo bias, *would*, in fact, choose the machine. This counterargument has been put forward by the psychologist Felipe de Brigard<sup>3</sup>. He suggests a way to pose the experiment to counter the notion that people wouldn't choose the machine, is by putting forward the idea that we have already chosen to enter the machine in the past (multiple times), and that, should we feel differently about this version of the experiment, it would be due to status-quo bias.

Another reason why Nozick is wrong, is due to the fact that this thought experiment relies upon the idea that either A) scientists are able to quantify and provide value to true human

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Experience\\_machine](https://en.wikipedia.org/wiki/Experience_machine)

<sup>2</sup> Nozick, Robert (1974). *Anarchy, state, and utopia*. New York: Basic Books. pp. 42–45. [ISBN 0-465-09720-0](#)

<sup>3</sup> Katarzyna de Lazari-Radek; [Peter Singer](#) (22 May 2014). *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. OUP Oxford. p. 257. [ISBN 978-0-19-102242-5](#).

happiness, translate that into computerized information, and create a machine processing human feelings and emotion and adapting and intaking it, which then leads to a scenario where happiness as defined in the experiment has no true value, meaning either utilitarianism ceases to exist, or the utilitarian definition of happiness is different enough that it is transcendent of any physical format, or B) that happiness within a machine or isolated from reality still counts as happiness and/or utility. By separating happiness from reality, Nozick places the experiment in a position where it is unable to challenge utilitarianism due to the fact that utilitarianism relies on happiness grounded in reality, while the experience machine relies on happiness grounded outside of the confines of the physical world of the observer of said experiences. In attempting to criticize and counter hedonistic utilitarianism with an experiment that relies on the subject being separated from the physical realm, all Nozick does is support the idea that utilitarians are correct. Humans would not plug in due to the fact that to achieve maximum utility/happiness for all parties, all parties must exist and be present in reality first. To isolate ones self from reality, the amount of happiness they could experience in a machine comes at the price of the sum of all the happiness they, and all other human beings they could possibly interact with, could experience at once. The exponential sum of missed, true utility, invalidates any amount of personal, faux utility one could “experience”. To attempt to criticize an ethical theory grounded in the presence of the subject and all other subjects, using a thought experiment that requires isolating the subject from the reality the theory requires to be present in, is absurd, and the reason why Nozick is wrong. It is similar to questioning if a commercial pilot provides more utility in a flight simulator, or in an in-flight plane. The key component of reality is missing from one, just as it is missing from Nozick’s thought experiment.

P.S: In this writer's opinion, none of this matters, as hedonistic based utilitarianism is incorrect, is based off of a definition of happiness and utility that does not translate to what is expressed through art based off of the human experience, and due to the fact that at the end of the day, nihilism has already won and human ethics will continue to be flawed until we transcend our animal nature in some manner and place the collective over the one (Nietzsche's Ubermensch is one way in which this might occur, although it's quite optimistic to think that one person could deliver humanity to the next stage of being).