

Mirjana Sokić*

IS HAPPINESS IN THE HEAD?

Abstract: This paper examines the philosophical implications of Nozick's thought experiment, specifically focusing on the assumption that most people would not want to be plugged into the experience machine. I present an "inverted" experience machine scenario in order to argue that this assumption is incorrect and that the scenario raises important philosophical questions about our purported unwillingness to be plugged in. The paper concludes that the "inverted" experience machine scenario is compatible with the central thesis of hedonism and other internalist theories of well-being, and provides strong support for the idea that happiness is truly in the head.

Keywords: Nozick's thought experiment, anti-hedonistic argument, experience machine, internalist theories of well-being, subjective aspect of experience

1. Introduction

Robert Nozick's experience machine thought experiment, which appears in his book *Anarchy, State and Utopia* (1974), was originally intended to make a point about the morally unacceptable treatment of animals (Weijers 2011a).¹ However, shortly after the book was published, many philosophers took Nozick's thought experiment as one of the strongest objections to hedonism, and possibly to all positions that view our well-being or welfare as exclusively dependent on the *subjective aspect* of experience. According to this popular opinion, Nozick's thought experiment

* The Institute for Philosophy, Faculty of Philosophy University of Belgrade, mirjana-sokic19@gmail.com

1 Nozick argues that until we can explain why most people would not want to be plugged into the experience machine and show that the same reasons do not apply to animals, we cannot claim that only the conscious experiences of animals determine the limits of acceptable behaviour towards them (Nozick 1974: 43). His point is intended to refute the common argument against ethical veganism, which suggests that an animal's pleasant life justifies its mutilation, restriction of freedom, and eventual killing. For a defence of this anti-vegan argument, see Zangwill (2021).

raises important questions about the nature of happiness, the value of our experiences, and the limits of hedonism.² It suggests that many people, despite the allure of the experience machine, would not choose to be plugged in, because they value something more than their own subjective experiences.³ This challenges the hedonist view that all that matters is how our lives feel from the inside, and implies that there are other factors that contribute to our well-being. To better understand the philosophical implications of this thought experiment, let us consider it in its entirety:

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your next two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. Others can also plug in to have the experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everyone plugs in.) Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* (Nozick 1974: 42–43)⁴

As we can see, Nozick describes a fantastic scenario in which the reader is asked to imagine that they are given the choice to be plugged into an “experience machine” that would provide them with any experiences they desire, without any negative consequences or limitations. The machine would allow them to live out their wildest dreams, have the most thrilling experiences, and achieve all their goals, without any effort or risk. The only

2 The term “hedonism” is somewhat ambiguous (see, Weijers 2011b), but the experience machine thought experiment is typically used to challenge the generic version of this view, which holds that pleasure (*hēdonē* [ἡδονή]) and only pleasure intrinsically contributes positively to well-being (Weijers 2014: 514). For more information on other objections that have been raised against hedonism, see Shafer-Landau (2018).

3 In the relevant literature, positions that hold that subjective experience is the only thing that has intrinsic value are often referred to as “the internalist mental state theories of well-being” (Weijers 2014).

4 For an updated version of this scenario, see Nozick (1989: 104).

catch is that all their experiences would be artificially created and simulated. Nozick asks the reader whether they would choose to be plugged in or prefer to live a “real” life, with all its challenges and hardships.

Nozick’s answer to the questions at the end of the paragraph is that people’s intuitive reaction to the presented dilemma would be to stay unplugged (see, De Brigard 2010: 43), and that, more importantly, the fact that the vast majority of people would want to stay in reality – even though this reality might turn out to be less pleasurable for them – shows that, contrary to the central thesis of hedonism, there are other things that matter to us in addition to our experiences or, to use Nozick’s phrase, how our lives feel from the inside. Yet, despite the influence that the quoted paragraph has on the philosophical debate about the plausibility of hedonism, numerous authors now reject Nozick’s conclusion (see, Linn 2016: 315–316). Thus, for example, Harriet Baber (2008: 133–8), Matthew Silverstein (2000: 279–300), Jason Kawall (1999: 381–87), Sharon Hewitt (2010: 331–49), Alex Barber (2011: 271), Torbjörn Tännsjö (2007), and many others, maintain that Nozick’s thought experiment does not make a compelling case against hedonism. The primary goal of this paper is to provide insight into how to resolve this debate.

I will introduce and analyse recent formulations of Nozick’s anti-hedonistic argument in the relevant philosophical literature. This analysis will reveal that the main weakness of the thought experiment is the assumption that most, if not all, people would not want to plug into the experience machine. By presenting the “inverted” experience machine scenario, I will demonstrate that this assumption is incorrect. In the conclusion, I will argue that the “inverted” experience machine scenario raises important philosophical questions about our purported unwillingness to plug into the experience machine, and that it is compatible with the central thesis of hedonism and other internalist theories of well-being that view the subjective aspect of experience as the only thing with intrinsic value. I will thus conclude that the “inverted” experience machine scenario provides strong support for the idea that happiness is truly in the head.

2. Formulations of Nozick’s argument against hedonism

As previously mentioned, many philosophers believe that Nozick’s thought experiment provides a powerful and persuasive argument against hedonism and other internalist theories of well-being. However, it should be noted that pleasure and well-being were not explicitly mentioned in

the original formulation of the scenario. This means that in order for the thought experiment to refute hedonism, it must be restated appropriately. In this section, I will examine some versions of Nozick's anti-hedonistic argument that have been put forward in recent literature. One simple version of the argument is presented by Ben Bramble (2016: 137):

- [A] Plugging in would not be best for one.
- [B] Hedonism entails that plugging in would be best for one.
- [C] Therefore, hedonism is false.

In order to see where this argument goes wrong, we should recall that, strictly speaking, Nozick nowhere suggests that plugging into the experience machine 'would not be best for one' – as is stated in premise [A] – but rather makes an *assumption* that the intuitive reaction of most people would be to stay in reality, even if plugging in would turn out to be significantly more pleasurable. As Bramble explains, a possible rationale for premise [A] emerges from the following reasoning: the psychological fact that, given the option, most if not all people would decide not to plug into the experience machine represents the most reliable indicator that plugging in would not be best for them (see, Bramble 2016: 137). Notwithstanding the initial plausibility of this reasoning, the point to keep in mind is that to identify the content of people's wants or choices with what is good for them is problematic to say the least (see e.g., Kawall 1999; Silverstein 2000; Hewitt 2010). At any rate, without some additional and conceptually independent elaboration on how we can identify *what is desired with what is desirable*, the question-begging character of this argument is evident (see, Baber 2010). Given this difficulty, I think we should sidestep the entire debate about whether Nozick, in fact, incorporates a claim similar to premise [A] in the original version of his scenario, by considering what I take to be a significantly superior formulation of Nozick's anti-hedonistic argument. The formulation in question was proposed by Dan Weijers (2011a: 229–231), and it runs as follows:

1. Plugging into an Experience Machine would make the rest of your life dramatically more pleasurable and less painful than it would otherwise have been (stipulated in thought experiment).
2. Given the choice to plug into an Experience Machine for the rest of your life, ignoring any responsibilities you might have to others, you would decline (appeal to readers' judgment).
3. If, ignoring any responsibilities you might have to others, you would decline the chance to plug into an Experience Machine for the rest of your life, then pleasure and pain are not the only things of intrinsic value (or disvalue) in a life.

4. Pleasure and pain are not the only things of intrinsic value (or disvalue) in a life (modus ponens, [2], [3]).
5. If hedonism is true, then pleasure and pain are the only things of intrinsic value (or disvalue) in a life.
6. Hedonism is false (modus tollens, [4], [5]).

A great deal could be said about this argument. First, there is no doubt that it is valid, for the premises adequately support the conclusion. Moreover, observe that premise [3] successfully avoids all of the problems that we have encountered with the previous argument; namely, it does not say that plugging in is *not best* for people; rather, it is a conditional according to which if it truly is the case that, when presented with Nozick's scenario, people would decide not to plug in – i.e., if they would choose not to abandon their current life in reality in favour of a much more pleasant virtual life – then pleasure could not be the *only* thing that is valuable in itself, as hedonists argue. The question that remains here, on the other hand, is whether the antecedent is true. In the following section, I intend to examine if Nozick was correct in assuming that most people would refuse to plug into the experience machine.

3. “Real” pleasure and “illusory” displeasure

As we have seen, the biggest challenge with Nozick's anti-hedonistic argument is the assumption that people's intuitive response to the opportunity to plug into the experience machine would be to stay in reality. Some philosophers believe that this response is due to confusions about the concept of reality, as well as misunderstandings about the implications of the perfect illusion that the experience machine is supposed to create. However, most authors argue that the negative response to Nozick's scenario is due to the so-called *status quo bias*, which is “an inappropriate preference for things to remain the same” (Weijers 2014: 530; see also Bostrom & Ord 2006; De Brigard 2010: 44). This bias typically manifests as a preference for the source of one's experiences to remain the same, regardless of whether those experiences are virtual or real. To overcome the problem posed by the status quo bias, Weijers created a scenario in which neither reality nor the experience machine is presented as the status quo (Weijers 2014: 252). In addition, unlike in the original Nozick scenario, the purpose of Weijers's version of the experience machine thought experiment is to determine people's intuition about whether it would be best for *someone else*, named Boris, to plug into the experience machine for the rest of their life. Interestingly, Weijers found that 55% of 77 partici-

pants said that the best option for Boris would be to plug into the experience machine for the rest of his life. These results indicate that Nozick's assumption in premise [2] – that the vast majority of people would prefer to remain in reality – is factually incorrect.

Weijers's scenario with Boris, while involving a decision about someone else rather than oneself, shares an important similarity with Nozick's scenario. Namely, Eden Lin (2015) notes that one of the most problematic aspects of Nozick's scenarios is that they typically involve a life within the experience machine that is hedonically superior to real life. In contrast, Lin proposes testing hedonism by considering two lives that are "experientially and thus hedonically identical" (2015: 320). Lin asks us to compare the life of Adam, who lives in the real world, with the life of Bill, who was plugged into an experience machine immediately after birth. Lin stipulates that the lives of Adam and Bill are identical "with respect to the qualitative features, durations, and temporal distribution of the pleasures and pains they contain" (2015: 321), which should mean that they are equal in welfare as well. However, when comparing these two lives, we may feel that there is something pitiful about Bill's life, but not Adam's. This suggests that Bill's life is somehow lower in welfare than Adam's, despite being hedonically identical. In short, Lin's example shows that if we feel bad for Bill, it can only be because Adam's life is higher in total welfare than Bill's, and since their lives contain the same amount of pleasure, we can conclude that the central thesis of hedonism must be incorrect. Yet, it is possible to object that many people believe that there is something pitiful about Bill's life because they have a conceptual confusion and bias towards the concept of "reality". They tend to believe that if something is real, it is *more valuable* than something that is virtual, even if the virtual thing provides the same experiences. This line of thinking becomes clear when we ask *why* people would feel bad for someone who has a life filled with pleasant experiences, but those experiences are part of a perfect simulation within an experience machine. The answer, that the experiences are not real and therefore less valuable, only begs the question against hedonism. So, where do people get the idea that a perfect illusion or simulation of reality is in any way worse than a real experience that is indistinguishable from the illusion?

The answer to this question can be found in Bart Engelen's interesting paper, in which he discusses the philosophical implications of Nozick's thought experiment and uses the film *Open Your Eyes* (*Abre los Ojos* 1997), directed by Alejandro Amenábar, as a reference. Engelen provides a thorough and insightful analysis of the thought experiment and its relevance to debates about the nature of reality and our experiences. The movie tells the story of César, a wealthy and handsome young man who

is disfigured in a car crash. He has a series of horrifying experiences and eventually learns that he committed suicide after the crash, but signed a contract with *Life Extension* to be cryogenically preserved until technology could revive him and attach him to a machine that would replace some of his memories. This machine, just like Nozick's experience machine, would allow him to live a virtual life of his choosing. However, César's machine is not functioning properly, leading to a nightmare-like existence.⁵ The only way to disconnect from the machine is to commit suicide, which César eventually does.

In addition to its impressive cinematic qualities, *Open Your Eyes* raises serious questions relevant to Nozick's thought experiment. Engelen points out that the perfect illusion created by the experience machine is indistinguishable from reality. This leads us to ask: are someone's horrifying experiences less dreadful and disturbing because they are not happening independently or outside of their experiential perspective? A modified version of Lin's example with Adam and Bill can help answer this question. Suppose that Adam and Bill both have lives filled with horrors, tragedies, and unpleasantness, and the only difference is that Adam lives in reality while Bill's conscious experiences are the result of being plugged into the experience machine. In this case, we would feel bad for Adam, but the key question here is whether Bill's life would warrant *less* pity since his experiences are not real. I admit that it is notoriously difficult to provide a definitive answer to the question of whether the person connected to the experience machine, inside of which she suffers from *virtual* depression, is better off than the person suffering from depression that has *natural* or *real* causes. However, based on the fact that the perfect illusion is indistinguishable from reality, I am inclined to think that the answer to this question is negative.

To further support this point, let us consider a scenario in which two people, person *A* and person *B*, experience the same tragic event. After the tragedy, person *A* passes away without any additional complications, while person *B* suffers from severe depression. Despite all objective factors and circumstances being the same for both people, it is hard to deny that person *B*'s life is worse than person *A*'s, as depression is a factor that we take into account when evaluating the quality of someone's life, even though its effects are limited to the individual's subjective experience. Now, consider a hypothetical individual, *C*, who experiences the same situation as individuals *A* and *B*, but who is immediately connected to an experience machine that provides her with the same experiences as individual *B*, with the only difference being that person *B*'s depression

5 Engelen notes that Amenábar wrote the script for the film *Open Your Eyes* after experiencing a series of unpleasant hallucinations due to a high fever (Engelen 2010: 44, note 1).

was natural, while person *C*'s was the result of the experience machine. Is it accurate to assert that person *C*'s circumstances are "better" – in the sense of deserving less pity – than those of person *B*? I personally tend to answer this question with a resounding no, but it would be interesting to see if most people's common-sense intuitions agree with this answer. This thought experiment illustrates how our evaluation of the quality of someone's life is not solely based on objective factors, but also takes into account the individual's subjective experiences.

4. The reversed scenario: Neo's dilemma

Adam Kolber's thought experiment modifies one aspect of Nozick's scenario while keeping the other aspects consistent and focusing on the same issue (Kolber 1994: 15). In his scenario, the reader is already hooked up to an experience machine and is asked if they would like to remain connected or go to reality. Kolber argues that more people would choose to stay connected to the experience machine in his reversed scenario than would agree to be hooked up in Nozick's original scenario (1994: 15). This thought experiment explores the implications of being in an experience machine and the choices we might make if given the option to remain in it or leave. One way to present this reversed scenario is as follows:

Imagine that you are currently hooked up to an experience machine. All the beings you have interacted with so far, including your family, friends, acquaintances, and pets, are part of the perfect illusion created by the machine. Your entire life, which you thought was real, is actually just a carefully designed program. If you disconnect from the machine, you will meet *real* people, form *real* friendships, find *real* partners and pets, and so on. However, you have been warned that if you disconnect from the machine, you will permanently lose contact with all the people and things you believed to be real while you were hooked up. Given this information, *would you choose to disconnect from the experience machine?*

This thought experiment raises questions about the nature of reality, our relationships and experiences, and the value we place on them. It also challenges our assumptions about what it means to be "real" and whether the reality we perceive is the only one that matters. The dilemma faced by Neo in *The Matrix* (1999) is similar to the dilemma presented in the above hypothetical scenario. Neo finds out that his previous life was an illusion created by a computer program, and he must choose between taking the red pill and leaving the illusory world for a real life, or taking the blue pill and continuing to live in ignorance in the illusory world. In the movie, Neo's life is shown to be very unfulfilling, and there is no information

about his relationships with others. This presentation of his life justifies his decision to take the red pill, which is consistent with Nozick's conclusion.

However, it is worth considering whether we would act like Neo in this situation. In other words, with Engelen's assertion that the perfect illusion is *phenomenologically indistinguishable* from reality in mind, it is questionable whether we would choose to abandon our previous lives (including friends, family, partners, and pets) if we were told that these entities do not actually exist independently of the experiences created in us by a computer program. This thought experiment challenges us to consider the extent to which our relationships and experiences are valuable to us, and whether we would be willing to give them up for the sake of "reality". This raises further questions about the value and meaning of our relationships, experiences, and emotions, and whether they are ultimately based on a real or an illusory reality. It also prompts us to think about what it means to be "real" and whether the *reality that we perceive* is the only one that matters. In Nozick's thought experiment, the reader is presented with the dilemma of choosing between a life in reality or a life in a perfect illusion created by an experience machine. While Nozick argues that most people would choose to remain in reality, we have seen that many philosophers have criticized this assumption and pointed out that people's decisions in such scenarios may be influenced by irrational factors, such as the desire to maintain the status quo and the continuity of their experiential identity. Despite the potential problems with Nozick's thought experiment, it still raises important philosophical questions about the value of our experiential perspective and the role of reality in determining the quality of our lives.

5. Concluding remarks

The results of this paper indicate that our understanding of the concept of reality and the role of our experiential perspective is fraught with confusing intuitions.⁶ As technology continues to advance and the devel-

6 Philosophical discussions often involve confused intuitions about the concept of reality and the role of our experiential perspective. For instance, Derek Parfit's (1984) thought experiment with teleportation challenges our everyday understanding of personal identity. Imagine a situation in which scientists have developed a teleportation machine that can scan a person's entire body (destroying it in the process), transmit the information to a distant location (such as Mars), and recreate a new, qualitatively identical body from the same particles. The question posed is whether this procedure allows a person to travel at the speed of light, or whether it simply kills one person and creates another, qualitatively identical one at the destination.

opment of virtual reality becomes increasingly possible, scenarios like Nozick's thought experiment become not only a theoretical possibility but a potential reality. In order to properly evaluate the value of our experiential perspective in relation to what is considered "real", it is necessary to examine how variations of the experience machine thought experiment affect human intuition and determine whether these intuitions are influenced by irrational or extraneous factors.⁷ In conclusion, my final answer to the central question of this paper – "Is happiness in the head?" – is that happiness appears to be in the head, at least in the sense that people's experiential perspectives and conscious experiences play a crucial role in determining the quality of their lives and the overall amount of welfare. This answer aligns with the fact that no one would agree that a person leads a happy life based solely on objective circumstances, while ignoring the person's own subjective experiences and overall situation. This is true regardless of whether those experiences are the product of an experience machine or are real.

Also, we have seen that Nozick's assumption that people would have a negative intuitive response to being plugged into an experience machine is largely accurate, but only in the sense that, due to various irrational factors such as conceptual confusions, irrational fears, and the status quo bias, people do not always choose happiness. While Nozick's scenario may seem to be against hedonism, the question remains whether it poses a conclusive challenge to the philosophy, considering that the most common reason people give for refusing to plug into the experience machine is based on irrational and irrelevant considerations, as well as confusion regarding key concepts such as *happiness* and *reality*. It is, thus, reasonable to adopt Feldman's conclusion that, even if most people would refuse to plug into the experience machine, Nozick's position against hedonism and other internalist theories of well-being does not hold (see, Feldman 2011: 67–70). While I recognize that such a "hybrid" solution – which attempts to reconcile various viewpoints and theses (despite their popularity and increasing prevalence in contemporary philosophical literature) – is not

Many people tend to view the described procedure as a kind of advanced killing, distinct from regular killing in that it creates a person who is qualitatively the same as the one who was killed. Even if we are confident that the teleportation machine will work perfectly, would we be willing to let our loved ones "travel" in this way, knowing that their original body will be destroyed and replaced with a numerically different one? Parfit believes that any opposition to such a procedure is irrational and based on our prejudices regarding numerical identity.

7 Something similar is the case with the famous philosophical problem known as "the trolley problem". This problem has many different variants and formulations that are used to examine which factors determine our reactions. For more on this issue, see Edmonds (2014) and Kamm (2015).

particularly satisfactory, it is currently the only solution that appears to me to consistently and philosophically accurately consider all of the arguments and objections made in recent decades to Nozick's views in his well-known passage.

References:

- Baber, H. (2008). "The Experience Machine Deconstructed." *Philosophy in the Contemporary World* 15, 133–8.
- Bostrom, N., & Ord, T. (2006). "The Reversal Test: Eliminating Status Quo Bias in Applied Ethics." *Ethics* 116, 656–679.
- Bramble, B. (2016). "The Experience Machine." *Philosophy Compass* 11, 136–145.
- De Brigard, F. (2010). "If You Like It, Does It Matter If It's Real?" *Philosophical Psychology* 23, 43–57
- Edmonds, D. (2014). *Would You Kill the Fat Man?* Princeton University Press.
- Engelen, B. (2010). "Open Your Eyes? Why Nozick's Experience Machine Does Not Refute Hedonism." *Film and Philosophy* 14, 33–46.
- Feldman, F. (2011). "What We Learn from the Experience Machine." In Ralf M. Bader and John Meadowcroft (ed.), *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*. Cambridge University Press, 59–86.
- Hewitt, S. (2010). "What Do Our Intuitions about the Experience Machine Really Tell Us about Hedonism?" *Philosophical Studies* 151, 331–49.
- Kamm, F. M. (2015). *The Trolley Problem Mysteries*. Oxford University Press.
- Kawall, J. (1999). "The Experience Machine and Mental State Theories of Well-Being." *Journal of Value Inquiry* 33, 381–87.
- Kolber, A. J. (1994). "Mental Statism and the Experience Machine." *Bard Journal of Social Sciences* 3, 10–17.
- Lin, E. (2016). "How to Use the Experience Machine." *Utilitas* 28, 314–333.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Nozick, R. (1989). *The Examined Life*. New York: Simon and Schuster.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Shafer-Landau, R. (2018). *The Fundamentals of Ethics*. Oxford University Press.
- Silverstein, M. (2000). "In Defense of Happiness: A Defense of the Experience Machine." *Social Theory and Practice* 26, 279–300.
- Tännsjö, T. (2007). "Narrow Hedonism." *The Journal of Happiness Studies* 8, 79–98.
- Weijers, D. (2011a). "The Experience Machine Objection to Hedonism," In Michael Bruce & Steven Barbone (eds.), *Just the Arguments: 100 of the Most Important Arguments in Western Philosophy*. Wiley-Blackwell. 229–231.
- Weijers, D. (2011b). "Hedonism." *Internet Encyclopedia of Philosophy*. URL= <http://iep.utm.edu/hedonism/>
- Weijers, D. (2014). "Nozick's Experience Machine Is Dead, Long Live the Experience Machine!" *Philosophical Psychology* 27, 513–35.
- Zangwill, N. (2021). "Our Moral Duty to Eat Meat." *Journal of the American Philosophical Association* 7, 295–311.