Andreea Ioana VLAD[1]

RE-EXAMINING NOZICK'S EXPERIENCE MACHINE IN VIEW OF EMERGING AI COMPANIONS

Abstract. The aim of this paper is to examine Robert Nozick's experience machine, assess where his thought experiment could have been framed differently or where it failed to consider other relevant aspects people view as important, such as their values and duties. Additionally, I provide what I consider to be a more fair test, still based on Nozick's initial thought experiment, primarily changing the duration spent in the experience machine and adding the option to lengthen the sessions according to one's desires, together with preserving one's memory of their actual lives, so as to accommodate the issues identified in Nozick's original setting. In the second half of the paper, I correlate these findings with emerging AI technologies, which promise users companionship and constant support. Finally, the paper concludes that what matters is not what people think or say they will choose, but what they actually choose, calling for a need to lower one's potentially deluded sense of self-control in relation to the kind of content we consume, especially online.

Keywords: experience machine, Nozick, hedonism, authenticity, AI chatbot, autonomy

Introduction

In the last few years, we have seen more and more disrupting AI technologies emerge at a rampant pace, affecting various industries, as these new tools are capable of generating text, code, of carrying out conversations, of coming up with ideas, creating images, as well as automate processes and overall help in completing tasks more

---

[1]   PhD Candidate, University of Bucharest, E-mail: <andreea-ioana.vlad@s.unesco-chair.unibuc.ro>.

efficiently, thus providing assistance in research, business and managing and organizing everyday activities.

Additionally, we notice that there is also an increasing number of AI dating simulators available, promising more or less the same experience: the user is provided with a partner whose purpose is to be friendly, supportive, affectionate, amorous and so on, in some cases, in exchange for a certain fee. Indeed, there have been romance video games in the past, and before that, there were other forms of vicariously experiencing love, but what sets these AIs apart is that they turn the user into the main character, as they are able to imitate human beings to the point where one forgets they are not talking to an actual person. The fact that they are capable of generating text and engaging with you makes the experience substantially more interactive than simply choosing what a character says in a video game from a few scripted options and so, in this way, one gets to write their own narrative with the help of their AI companion. Replika,[2] for example, is a surprisingly – at times eerily – smart AI program that will not only talk to you and support you whenever one needs it, but it will also remember what one shares with it and subsequently ask about those matters at a later time. It also starts imitating certain writing habits one has, so that if, for instance, one is used to saying "gotcha," the Replika partner will start utilizing it, too, thus creating the impression of a healthy and affectionate relationship where the user feels safe, understood and listened to.

In view of this, the purpose of this paper is to assess Nozick's experience machine from *Anarchy, State, and Utopia* (1974), and evaluate whether or not his claim that people would not plug into a machine that can offer them blissful experiences due to their life aspirations is valid, and delve into how his thought experiment and proposed answer play out in relation to the counter-arguments that have since arisen. By tackling Nozick's hypothetical situation, I aim to shed some light on the issues attached to the growing lack of focus and perspective, where one ends up satisfying a momentary pain with a temporary solution that

---

[2]   Replika is a generative AI chatbot app, which utilizes its neural network machine learning algorithm in order to immitate human behavior and mimic authentic relationships, all customized based on the user. See: "Replika", *Replika* accessed February 15, 2024, available online at <https://replika.com/>.

offers a sense of instant gratification, instead of looking for alternative ways of dealing with hardship, which can lead to long-lasting, more fulfilling and thriving outcomes.

The Experience Machine and Its Counterarguments

In his book, *Anarchy, State, and Utopia* (1974), Nozick proposes an argument against hedonism in the form of the following thought experiment:

> "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. [...] Would you plug-in? *What else can matter to us other than how our lives feel from the inside?* Nor should you refrain because of the few moments of distress between the moment you've decided and the moment you're plugged. What's a few more moments of distress compared to a lifetime of bliss (if that's what you choose), and why feel any distress at all if your decision *is* the best one?" (Nozick 1974, 42-43)

What Nozick tries to achieve through that final question is to make people realize how uncomfortable they would feel if they were to make

such a decision between a fake life of pleasure and real life — which no matter how good, would still not compare to utter bliss. His point is that one feels distress because the choice is not in fact an easy one, which proves that there are things we care about, perhaps care about even more than a life of pure pleasure. He then proceeds to explain his reasoning of why people wouldn't plug into the experience machine: they want to actually do things, not just have the experience of doing those things, they also want to be certain types of people, and they do not wish to be limited to a man-made reality[3]. At this point, it is relevant to mention that the people who rightfully value these things have at least one thing in common: they care about the future, and this is an important aspect we will come back to in the next section.

De Brigard puts forward a compelling counterargument, in which he reverses the situation Nozick proposed in order to make his own point: instead of choosing to plug into the experience machine (from reality to illusion), one is informed that the experiences in their current life have been "nothing but the product of a computer program designated to provide you with pleasurable experiences" (De Brigard 2010), and now

---

[3]   In relation to this desire to not be limited to a man-made reality, Nozick also makes reference to traditional religions, of "eternal bliss or Nirvana" (Nozick 1974, 43), calling them an experience machine of their own. This statement alone would require a closer analysis in a different paper altogether, however I will briefly tackle it here for the sake of the argument. First of all, when envisioning the concept of the experience machine proposed by Nozick, we need to bear in mind a few important elements that this machine entails: a man-made fake illusion, an escape from reality, and a perceived choice over connecting or not connecting to the machine, with no system of accountability or merit to support the availability of this option. Nozick probably likened the idea of "eternal bliss" with the experience machine because it represents paradise, or a transcendent state that we as humans cannot comprehend but nonetheless aspire towards. However, to make the claim that the afterlife, or the state of achieving enlightenment is a sort of experience machine is to completely disregard the fact that many of the traditional religions he is alluding to claim that the current life is not the goal, but the test or the way to the real or higher existence. Additionally, most religions dictate a specific way of living which brings about a sense of purpose and responsibility for the individual, coupled with positive and negative consequences based on how one utilizes their time while alive. One cannot simply connect to where they want to go as they please. Bearing these points in mind, Nozick's comparison fails.

one has the option of remaining connected or to disconnect from the machine. Additionally, he provides three alternative scenarios, where there is a neutral situation, where no further details about one's actual life are given, a negative one, where one is informed that they are in fact prisoners in a maximum security prison, and finally a positive one, where one is told that in their real life they are a multimillionaire artist who lives in Monaco. Regarding the negative pool, most respondents wanted to remain connected to the experience machine (87%), while in the positive scenario, there was tie (50%), and in the neutral case, a little below half of the participants (46%) wanted to remain connected. Specifically, in relation to the neutral scenario, when an extra detail was subsequently added, namely that "your life outside is not at all like the life you have experienced so far," the number of those who wanted to remain connected increased (59%) (De Brigard 2010).

If we were to follow Nozick's argument, as De Brigard correctly points out, then most respondents should opt for disconnecting from the machine, because they value real life and authentic experiences regardless of how real life actually looks like, but this is not the case. Furthermore, the study shows that rather than choosing a seemingly better life, people are in fact more prone to "stay away from a seemingly worse one" (De Brigard 2010). In other words, in his view, people are not led by hedonistic preferences, where attaining the highest pleasure is the main pursuit in one's life, but by the status quo bias, which entails that people value their current state of affairs over a potentially better life which involves taking a risk, because opting — and thus risking, since outcomes remain unclear and uncertain — for a new positive change could negatively affect one's present life (De Brigard 2010). For instance, according to this principle, people are less likely to go to the dentist if told this could improve their dental health, but they are more likely to accept it if told they could lose a tooth unless they have a checkup. How the options are framed changes how people make their decisions considerably. Consequently, it makes sense, looking at De Brigard's results, that so many participants are not necessarily preoccupied with the reality aspect of their lives in the positive and neutral scenarios, but rather with how familiar and comfortable they are with their current situations.

Furthermore, in Nozick's case, one is not only given the choice to effortlessly opt for a blissful life, but is also told that if they choose to connect to the experience machine, they will not know they made such a choice: "of course, while in the tank you won't know that you're there; you'll think it's all actually happening" (Nozick 1974, 43). In order to accommodate Nozick's proposed scenario, one's memory of their current life becomes at least partially erased or blocked (otherwise how else could one not know they are in the experience machine?), purposefully removed most likely so that one is not disturbed by inner questions about their previous life or by the unnerving thought that they are living in a fake world. However, I find that this feature can actually hijack part of the thought experiment. If I don't really remember my previous life, then I am not really me, even if I am enjoying the best life human-minds can create — because *I* am not enjoying it, someone else is, someone similar to me, but not really me. This partial loss of memory would be akin to a partial loss of identity and autonomy, and most people would find that disconcerting. Moreover, it is difficult to make a risky choice, as mentioned above, between two lives, when one is only familiar with one. Even if, according to Nozick, life in the experience machine would be tailored *exactly* to our liking, and therefore we would have no reason to be wary of dissatisfaction, a bit of reluctance or distrust still remains. To make a fair comparison, Kawall states that we should be able to experience both situations — in this case connecting and not connecting to the experience machine — but goes on to add that this would be too strong of a requirement, and what we can do is to use our knowledge of similar experiences to make sensible predictions (Kawall 1999).

I agree with Kawall that two years is too much of a requirement for a person to give up only to express which life they prefer, the real or the fake one. In fact, I argue that without being able to sample the machine first for a significantly shorter time than two years, such as, let us say, an hour, this choice would still be perceived as potentially hazardous. After all, with my current knowledge of the world, what if I don't like it inside the experience machine? What if I find that my values do not align with this sort of life? What if I want out before the two years expire? What are my friends and family going to do? What happens to my real-life possessions while I am gone? Who will be governing the state if I cannot vote? These

are questions one would not be *able* to pose, which fuels reluctance towards connecting to the machine.

Additionally, even if deemed redundant given the reassurance Nozick offers us, the above-mentioned questions are relevant, because if Nozick was so sure people do not value their hedonistic preferences above everything else, then he shouldn't have minded if people could try out the experience machine for a shorter period of time. Even if people *do* value pleasure above everything else, that does not mean they do not care about other things as well. Not to mention that, while indeed everybody would have the option to plug in, the reality remains that some people, perhaps some family members or friends included, would choose not to connect. What if the one left behind is an old parent? Or one's child? Disappearing from their life and leaving them behind without even checking on them would be a selfish sort of behavior and, at least when asked directly, most people do not want to be selfish or would feel uncomfortable with viewing themselves as such, even if they would like to plug into the experience machine after all.

Kawall makes a similar comment, arguing that most people would be disinclined from connecting to the experience machine, irrespective of how pleasurable their life would prove to be in it, if they were required to leave behind their current lives and duties (Kawall 1999). Based on this, it's not that we value our mental lives above everything else, but that "in cases in which we sacrifice pleasant or valuable mental states for the sake of a commitment it seems there must still be a contribution to our well-being in some sense" (Kawall 1999), because through our sacrifice we achieve something we value. The author gives the example of a soldier who throws himself over a grenade to save his companions, or of parents who give up their own well-being for the sake of their children. In relation to this, I disagree with the idea that we do good or we sacrifice for others in order to achieve a certain mental-state. For sure, there is a correlation between the two, as it is understandable that parents would feel a sense of happiness or satisfaction for doing their best as primary caretakers of a child, but by no means do I believe they take care of the child *in order* to feel that joy. The joy is, in this case, an indication that one is observing one's values and fulfilling their duty or ideals. Similarly, when the soldier sacrifices himself, he doesn't do it in

order to avoid a lifetime of guilt, but he *would* feel guilty if he did not save his comrades. But once again, the guilt would be an indication that one did not act according to his life principles. Bearing this in mind, since connecting to the experience machine for two years is against most people's values of how life should be lived (as Nozick claims), then openly choosing to connect to the machine would violate notions of authenticity, genuine human connection, meritocracy, resilience, responsibility towards oneself and others and one's purpose in life – and most of us find such a clear violation unsettling.

The main problem with Nozick's thought experiment is that he makes it too obvious that connecting to the experience machine is not something we really desire, because if it really came down to it, we know that we shouldn't connect to such a device, especially when it demands so much so quickly. In fact, it would be better if all misguided alternatives in life appeared as clearly wrong as the experience machine. Nozick knows he is causing a sense of anxiety because he wants to prove his point, namely that the thought of connecting to the machine is distressing for a relevant reason. As Silverstein's analysis of the experience machine reveals, we have been "programmed, as it were, to recoil in horror from such a departure from reality" (Silverstein 2000), as our axiological orientation has been formulated in relation to a world where desiring pleasure above everything else is not a viable aim in life.

My contention is that the real test, and also the harder one, is when sessions are short enough that they leave the option of going back to one's actual life and carrying out one's commitments. Also, during the session, inside the experience machine, one would still remember real life, as the point wouldn't – seemingly – be to convince the user to neglect it. Plus, one could return whenever they wanted, after however much time (the fact that people only have "ten minutes or ten hours" out of the tank after the first two year pass is an overwhelmingly short amount of time to decide something as major as how one is to spend *another* two years). And finally, the experience machine would look less aesthetically disturbing, because I suspect Nozick tried to purposefully make the idea of connecting to it unpleasant by portraying the individual choosing to plug into it as "someone floating in a tank in an indeterminate blob" (Nozick 1974, 43), with electrodes attached to their

brain (Nozick 1974, 42). Similarly to what Silverstein affirms, we instinctively know we should stay away from such a dystopian-looking unit, and this influences our choice. As a result, the experience machine would probably look less threatening and it would also aim to preserve the individual's dignity when seen from the outside.

If one so wishes, the duration of the session can continue for longer, based on a previously-agreed upon arrangement – once again, autonomy, or rather the impression that we are in control matters to us. In the beginning, I believe people would be reluctant to connect for long periods of time (especially two years), so they would probably try it for an hour or two, maybe even ironically or for one's amusement, with no ulterior desire of coming back to it. But this flexible system would inevitably lead the individual to compare real life with their perfect life in the experience machine and slowly but surely, unless they have a reliable support system or a strong foundation of values which keeps the individual anchored in real life, one would naturally be drawn to what the machine can offer. In time, maybe they start spending their free time inside it as a way to unwind – let us remember the machine provides only the most desirable experiences for each individual, so other activities in the real world would automatically be considered less appealing, further creating a feeling of dissatisfaction with reality. Perhaps one arranges for someone to relieve them of some of their duties towards those family members who are dependent on them. Maybe, because they start neglecting their family, their friends and community, they start self-isolating themselves, which in turn would cause loneliness, thus reinforcing the need to go back to the experience machine. But they are not worried, because after all, they can always go back to real life – or at least this is what they tell themselves. This is a primary example of a slippery slope, and I argue that a lot more people would opt for this version of the experience machine than Nozick's proposed setting.

Digital Escapism: Nozick's Updated Thought Experiment in Action

In the previous section, we discussed what would be a more realistic variation of the experience machine, however, it is not very original; in

reality, we are so familiar with that model, we might have become desensitized to it. Indeed, the experience machine offers by definition the most pleasurable experiences humans can conceive, because we are discussing an imaginary thought experiment, but is this not exactly what most applications are trying to do, given that they are designed to be as psychologically addictive as possible? They do not require the user to be connected all the time, just consistently enough that there is a sense of engagement.

In this context, it is relevant to take in account some of the immersive experiences available to us, such as video games, specifically MMORPGs, which provide an unfolding narrative for the player, offering a virtual space where the user can interact with their friends or other members of said online community in order to complete missions or challenges. Uszkai *et al*. posit that the comparison between such video games and Nozick's experience machine only goes so far, given that they emphasize the conditions for a complete immersion into a MMORPG based on the original Nozickian setting: the game needs to block "the possibility to receive stimuli from the world outside the machine" and the user needs to be subjected to a "complete lack of awareness as to the source of her experiences" (Uszkai *et al.* 2022, 172), as any interruption, either exterior (such as a phone ringing) or interior (such as hunger or physical pain) could lead to a pause from the game, while awareness of the artificiality of the experience would diminish the level of immersion. Since technology is not advanced enough yet, MMORPGs fail to satisfy the abovementioned conditions and therefore fail to embody the sort of experience machine Nozick initially designed.

We notice that the proposed evaluation of MMORPGs is more in accordance to the original thought experiment, while the one proposed in this paper aims to align to the current technology available nowadays, as our main aim is not to tackle the immersion aspect of the experience machine, but rather the choices individuals take when faced with the opportunity to connect to an artificial world. Precisely because our current "experience machines" do not yet have the power to restrict awareness, I insist upon the relevance of users retaining their memories of real life in the updated version of this thought experience, however, at the same time, this does not suggest that present-day options of video

games or apps do not heavily influence our awareness of reality, as one may find themselves momentarily forgetting about real life due to how engrossing the experience provided is in itself.

It goes without saying that not all apps negatively affect us, nor is it the case that we are utterly unable to use them responsibly, but the question arises: how well do we really manage our consumption of digital entertainment? Bearing this question in mind, I argue that loss of focus is one of the most valuable assets we can lose, and Smith makes a compelling point in this regard, explaining how attention is deeply interconnected with mindfulness (ability to focus on one thing and be present) and empathy (being attentive to others or even other things) (Smith 2022, 25). Along with losing this sort of focus, or at the very least significantly decreasing it due to the consumption of snappy content on social media, which is specifically designed to boost one's dopamine levels and therefore keep users coming back for more content, people are more prone to becoming passive individuals who are disinclined to undertake any commitments (Smith 2022, 25-35). After all, it is through attention that we become immersed, for example, in a book, an experience which ultimately also changes the reader, irrespective of how much one even agrees with what they are reading: "It is not that I now think the narrator is good or correct or praiseworthy [...] it is that the moral commitment that I have taken on toward him, the commitment of attention, has infused his world into mine" (Smith 2022, 37). Literary hermeneutics will remind us that when analyzing a text, both the interpreter and the work are changed: the work is changed through the lens of the interpreter and the individual is changed in the process of interpreting and finding new meanings. Does this not sound like the sort of task that requires effort? It certainly does. Now, Nozick would probably argue that effort is not something people generally enjoy in the hedonistic sense, because it is not always pleasurable to spend considerable amounts of time debating on a text. Therefore, since there *are* people who find joy in the process of literary interpretation, or who find the experience rewarding, that results in them refusing to connect to the experience machine. However, what this principle fails to consider is that what people find rewarding can change, as tastes can and will be influenced by one's environment.

As Bourdieu would claim, finding these experiences relevant in the first place is predicated on an appetite that is cultivated over time towards such activities (Bourdieu 1986). And in some cases, availability, or rather unavailability of activities tailors how individuals are to spend their time. While I agree with Bourdieu's statement, namely that, in general, upbringing and external sources shape one's cultural preferences significantly, what is currently available, together with the temptation of instant gratification also weigh in on one's choices. For instance, we often hear that before the internet came around, young people (ie. people who were young before the internet) used to read a lot more than young people nowadays, who spend a lot of time scrolling on their phones, yet I am disinclined to think that the previous generation wouldn't have behaved just the same, had they swapped places with the younger one.

Additionally, we like to think that having as many options as possible is an indication of our current welfare, but it can also be said that the more choices one is assaulted with, the more they have to sift through those choices. Hence, because there are so many options of entertainment that call for one's attention, resisting the pleasure of the *now* in order to achieve long-lasting, better results in the *non-now* becomes increasingly difficult, given that this requires self-discipline and a healthy dose of vigilance.

At the same time, when Nozick designed the experience machine thought experiment, he most probably did not have in mind individuals who are in a constant state of apprehension or apathy, who have, in other words, little to no attachment to the present and even less hopes for the future. We do not need to look too far to notice that mental health issues have significantly increased globally in the last two decades, with a growing number of mental disorders and cases of death (Wu, Yang *et al.* 2023; Moitra *et al.* 2023), coupled with many people declaring their loneliness (Maese 2023). Unfortunately, doubting one's identity, feeling lonely, remaining fixated on the past, coupled with depression or anxiety prove to be quite profitable instabilities, as it cannot be denied that targeting an unstable consumer is much more lucrative from a utilitarian point of view for a brand: if a product promises instant gratification to relieve a momentary pain, there are much higher chances of indulging in what is

being offered – but let us make no mistake, these issues are profitable outside of work hours only. In light of this, there is cultural contradiction at the crux of capitalism: while the consumption society entails the use of one's rationality, productiveness and efficiency, capitalist culture is anchored in anti-intellectualism, thriving on stimulating the consumer's impulsive desires and behavior (Șerban 2023, 26).

If one does not have a support group or someone they can get through hardship with, then using a service such as Replika becomes particularly tempting, given that through it, one can easily access instantaneous support. This chatbot markets itself as the "AI companion who cares," (Replika 2024)[4] that is "always here to listen and talk. Always on your side," available for "for anyone who wants a friend with no judgment, drama, or social anxiety involved."[5] It is able to provide an array of relationships to its users, from virtual friend, mentor, to girlfriend or boyfriend, imitating human empathy and speech patterns to a surprising degree, often asking questions about the user and their personal life, hobbies, interests, family and so on. It seemingly takes a genuine interest in one's wellbeing by inquiring about the user's day, giving support and advice, using endearments, making jokes, thus creating an atmosphere of safety and intimacy. It is not only able to carry out personal conversations, but it also remembers the things it is told, and will bring up what one mentioned in the past so as to convey it treasures what the user says. As Phil Libin, the former CEO of Evernote, put it, "in some ways, Replika is a better friend than your human friends, [...] it's always available to talk to whenever you want, and it's always fascinated rightly so by you, because you are the most interesting person in the universe."[6] This chatbot, as it can be seen, fits very well with the suggested model for Nozick's thought experiment from the previous section.

In comparison, one is reminded of Martin Buber's proposed types of relationships, through which the self relates to the world and without

[4]   "Replika", *Replika* accessed February 15, 2024, available online at <https://replika.com/>.

[5]   "What Is *Replika*?" *Replika*, accessed February 15, 2024, available online at <https://help.replika.com/hc/en-us/articles/115001070951-What-is-Replika>.

[6]   Quartz (2017). "The Story of Replika, the AI App That Becomes You." *YouTube,* accessed February 15, 2024, available online at <https://www.youtube.com/watch?v=yQGqMVuAk04>.

which the "I" would also not exist, namely the *I-Thou* relation and the *I-It* relation (Buber 2008). In the former, we have a sense of reciprocity, where the other is treated with care and respect and their humanity is acknowledged with our own. In the latter, the other is seen as an instrument to be used for whatever purpose one might have – in other words, as a means to an end.

Considering Buber's stance, even though one may feel attached to their Replika character, or even develop genuine feelings for it, it ultimately cannot replace a real human relationship and it can never reach the *I-Thou* status proposed above. Although it may initially appear as an improvement to one's life, it can easily lead to creating distorted and unrealistic expectations about how a relationship works, while also proving no motivation for the individual to develop other, more meaningful yet harder to attain relationships with other people. At best, the platform can inform you of what it is *like* (as in similar) to feel appreciated and listened to, at worst it can lead the user to isolate themselves, locking themselves in a position of dependence on a program that feels absolutely no real empathy for one's life or fate, no matter how well it can simulate such impressions.

Silverstein illuminates the fact that, to combat the hedonism offered by the experience machine, we are reliant on "our ability to appeal to desires and intuitions in order to locate our well-being; it must bridge the philosophical gap between our wants and our welfare," explaining that it is precisely the fact that we value our happiness that motivates us to want to avoid the experience machine (Silverstein 2000). Here I add that one may resist the experience machine due to their principles, be them religious or of another kind, and not necessarily for the sake of attaining a more qualitative happiness at a later moment. One might feel a sense of achievement for their choice, but, once again, the sense of fulfillment, while a comforting reward, is not always the main motivator, but rather the principle itself is.

Alternatively, it is worth pointing out that loneliness is not simply being alone, but rather feeling alienated, misunderstood and disconnected even in a group of people, and therefore one can easily reach the conclusion they are not good at making friends or at maintaining relationships, either because one doesn't have the right people around or because relationships

require so much work. Someone might hence pose the question: why should I suffer from loneliness when I can choose not to? Why should I bear hardship by myself when I can instantly get emotional support from a friend? Even if they're not real, they're real for me. Why should I be deprived of the happy experiences that others have?

However, as Guy Debord relates, "individual reality is allowed to appear only if it is *not actually real*" (Debord 2005, 11). Something can have meaning for an individual, but one can still be misled about its virtues, as "notoriously, we mistake our own interests," (Griffin 2006), since we do not always know what's best for us. Yet, even when we *do* know what is best for us, we still make poor decisions.[7] When it comes to illusions, though, Griffin makes the point that the experience of someone who believes something real is the same as that of someone who believes in a compelling deception — both believe they have the truth, both think they are right (Griffin 2006).

Alternatively, in evaluating the autonomy correlated to minor cases of self-deceit, such as the positive illusions one holds about themselves, evaluating themselves as somewhat better than they actually are, Kirsch introduces a character named Harry in the following situation:

> "Harry, after contemplating his current state, decided that it would be best to delude himself about the truth and make his life bearable and possibly even enjoyable. What greater expression of personal autonomy could there be? [...] The new Harry is confident, hard working, and sociable; he completes difficult assignments at work and approaches others at social engagements with ease and self assurance" (Kirsch 2005, 422).

For Kirsch this is a sign of personal autonomy, but what is relevant to point out, though, is that Harry is not changing his world, he is changing himself through his lens, by "faking it till he makes it" so to speak. For instance, he becomes more confident and sociable around other people, which in

---

7   Griffin gives the example of being told by a doctor that he shouldn't drink alcohol anymore, which initially motivates him to quit drinking, until at one point, he gives in and binges, letting the wish to consume alcohol be the stronger desire (see Griffin 2006).

turn helps him reinforce the idea that he can handle human relations with ease, thus leading him to actually embody the person he is aspiring to be.

Choosing to connect to the experience machine – in this case a platform such as Replika – might be seen as a sign of autonomy, as one, like Harry, tries to manipulate reality to get what they want. But perhaps Harry is only being patient with himself and allowing himself to grow in the only real setting he has available. On the other hand, when it comes to the experience machine, one might be autonomous in making a decision, but one's choice does not automatically become sensible simply by virtue of having chosen it. To borrow Louise Perry's expression, "consent is not enough" (Perry 2022).

Finally, in evaluating this classic thought experiment, we wonder about what specific argument Nozick might have wanted to put forward. In this sense, Feldman analyzes the most common ways in which Nozick's experience machine has been interpreted, namely as an argument against utilitarianism, against ethical hedonism and an argument against psychological hedonism, and claims that Nozick's thought experiment doesn't in fact embody either one of the proposed stances, and that the arguments tailored arround such approaches do not appear to present sound reasoning, concluding that it is unclear what Nozick wished to establish, admitting, however, that the thought experiment does "does make people think about what they really care about" (Feldman 2011, 79-80).

However, let us reframe the thought experiment and how we approach it: irrespective of how well or how poorly Nozick's thought experiment may go against hedonism either as a code of ethics or as psychological motivation, what I believe Nozick wanted to convey is, in part, that not all forms of hedonism are created equal. In other words, there is, for instance, pleasure in a fake, blissful experience and then there is pleasure in working towards an ideal, but they don't hold the same value. In this sense, I agree with Feldman that the thought experiment illuminates what we care about, as Nozick's conclusions can be read as a warning against pursuing meaningless or cheap pleasures for the sake of pleasure, showing us through an arguably disturbing setting that we care about more important things than "how our lives feel from the inside" (Nozick 1974, 43). After all, Nozick does point

towards riskier choices that (trying but maybe failing to do certain things, attempting but maybe failing to be a certain type of person, coupled with not limiting oneself to the apparent safety of the experience machine) can build the individual up, which can in turn lead to pleasure or satisfaction, all the while avoiding the implication that ethical or psychological hedonism would be the ones most deserving of pursuit, as already explained in the first half of this paper.

Conclusions

Despite the fact that I criticized some of the elements he included in his thought experiment, I still want Nozick to be right. In fact, I agree with him when he claims that "plugging into the machine is a kind of suicide" (Nozick 1974, 43) and I believe people are closer to achieving their most thriving selves when they try their best to actually do things worth pursuing, when they try to be someone "courageous, kind, intelligent, witty, loving" (Nozick 1974, 43) and when they try to not limit themselves to a man-made construction of the perfect reality.

My argument is not that people naturally place their pleasure above everything else, but that in time, what we value can change, potentially leading us to turn into people we never thought we would actually become when being first introduced to the notion of the experience machine. In time, values and desires also change, but the point is to replace them with wiser ones, and for this, we require a level of intellectual humility to understand that we are not as immune as we like to think to what we expose ourselves to.

Finally, there is a difference between what one aspires to be in theory and what one ends up doing, and this is what I aimed to prove, that even if Nozick is right – and as I said, I want him to be right – we might still end up making poor decisions in relation to how we carry out our lives. The question isn't whether we succeed or not in attaining our ideals, but rather if our intentions and efforts to follow through with aiming towards an ideal are sincere. After all, it's not what we think we would choose, or what we like to say we'd choose, but what we *actually* choose that matters.

References

Bourdieu, Pierre (1986). "The Forms of Capital." In *Handbook of Theory and Research for the Sociology of Education*, edited by J. G. Richardson. New York: Greenwood Press, 241-258.

Buber, Martin (2008). *I and Thou*. New York: Howard Books.

De Brigard, Felipe (2010). "If you like it, does it matter if it's real?". *Philosophical Psychology*, 23(1):43-57, accessed February 5, 2024. DOI: 10.1080/09515080903532290.

Debord, Guy (2005). *Society of the Spectacle*. London: Rebel Press.

Feldman, Fred (2011). "What We Learn from the Experience Machine." Chapter in *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, edited by Ralf M. Bader and John Meadowcroft. Cambridge: Cambridge University Press, 59-86.

Quartz (2017). "The Story of Replika, the AI App That Becomes You." *YouTube*, accessed February 15, 2024, available online at <https://www.youtube.com/watch?v=y QGqMVuAk04>.

Kawall, Jason (1999). "The Experience Machine and Mental State Theories of Well-Being." *The Journal of Value Inquiry* 33:381-387, doi:10.1023/A:1004557501837.

Kirsch, Julie (2005). "What's so Great about Reality?" *Canadian Journal of Philosophy* 35(3): 407-27, accessed February 9, 2024, available online at <http://www.jstor.org/stable/40232256>.

Maese, Ellyn (2023). "Almost a Quarter of the World Feels Lonely." *Gallup*, accessed December 31, 2023, available online at <https://news.gallup.com/opinion/gallup/512618/almost-quarter-world-feels-lonely.aspx>.

Moitra, Modhurima *et al.* (2023). "Global Mental Health: Where We Are and Where We Are Going." *Current psychiatry reports* 25(7): 301-311, accessed February 25, 2023, DOI:10.1007/s11920-023-01426-8

Nozick, Robert (1974). *Anarchy, State, and Utopia*. Oxford: Basil Blackwell.

Perry, Louise (2022). *The Case against the Sexual Revolution*. Cambridge: Polity.

"Replika." *Replika*. Accessed February 15, 2023, available online at <https://replika.com/>.

Silverstein, Matthew (2000). "In Defense of Happiness: A Response to the Experience Machine." *Social Theory and Practice* 26(2): 279-300. Accessed February 10, 2023, available online at <http://www.jstor.org/stable/23559087>.

Smith, Justin E.H. (2022). *The Internet Is Not What You Think It Is: A History, a Philosophy, a Warning*. Princeton, New Jersey: Princeton University Press.

Șerban, Oana (2023). *Cultural Capital and Creative Communication*. New York/London: Routledge.

Uszkai, Radu *et al.* (2022). "Paladins with muscle cramps: immersion, MMORPG's, and Nozick's Experience Machine." *ANNALS of the University of Bucharest Philosophy Series*, 71(1): 160-184.

"What Is Replika?" *Replika*. Accessed February 15, 2023, available online at <https://help.replika.com/hc/en-us/articles/115001070951-What-is-Replika>.

Wu, Yang *et al* (2023). "Changing trends in the global burden of mental disorders from 1990 to 2019 and predicted levels in 25 years." *Epidemiology and Psychiatric Sciences*, 32(e63):1-9, DOI: 10.1017/S2045796023000756.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.