The Experience Machine

Ben Bramble*
Lund University

Abstract

In this paper, I reconstruct Robert Nozick’s experience machine objection to hedonism about well-being. I then explain and briefly discuss the most important recent criticisms that have been made of it. Finally, I question the conventional wisdom that the experience machine, while it neatly disposes of hedonism, poses no problem for desire-based theories of well-being.

1. Introduction

Theories of well-being attempt to explain what it is in virtue of which lives can be good or bad for their subjects. According to one such theory, hedonism, lives can be good or bad for their subjects just in virtue of their ability to feel pleasure and pain (where ‘pain’ is shorthand for unpleasurable experience more generally).

Hedonism has a straightforward appeal. Many feel that something that has no effect on someone’s experiences does not ‘touch’ or ‘get to’ this person in the sort of way required for something to benefit or harm her. If this is true, then it seems only a relatively small step to hedonism.

Nonetheless, hedonism has few contemporary advocates. This is mainly due to a single, highly influential objection to it, widely considered to be decisive: Robert Nozick’s experience machine. Discussions of well-being – whether in scholarly journals, academic conferences, or university lecture halls – often begin with a quick dismissal of hedonism by reference to Nozick’s objection before turning to ‘more interesting matters’ (usually the question of which desire-based or hybrid theory of well-being is true).

In this paper, I will do three things: First, reconstruct Nozick’s objection. While the objection is oft-cited, it is rarely formulated in a clear or careful way. Second, explain and briefly discuss the most important recent criticisms that have been made of Nozick’s objection. Third, question the conventional wisdom that the experience machine, while it neatly disposes of hedonism, poses no problem for desire-based theories of well-being.

2. The Objection

Nozick’s most famous statement of the objection appears in his early work, Anarchy, State, and Utopia (1974). But it is his mature work, The Examined Life (1989), that contains his clearest formulation of it. There, he writes:

Imagine a machine that could give you any experience (or sequence of experiences) you might desire. When connected to this experience machine, you can have the experience of writing a great poem or bringing about world peace or loving someone and being loved in return. You can experience the felt pleasures of these things, how they feel “from the inside.” You can program your experiences for…the rest of your life. If your imagination is impoverished, you can use the library of suggestions extracted from biographies and enhanced by novelists and psychologists. You can live your fondest dreams “from the inside.” Would you choose to do this for the rest of your life?…Upon entering, you will not remember having done this; so no pleasures will get ruined by realizing they are machine-produced.
If hedonism were true, Nozick suggests, then ‘plugging in would constitute the very best life, or tie for being the best, because all that matters about a life is how it feels from the inside’.7

Intuitively, however, this is not so – there are alternatives that would be better for one. Therefore, hedonism is false.

We can state Nozick’s objection simply, as follows:

(1) Plugging in would not be best for one.
(2) Hedonism entails that plugging in would be best for one.

Therefore,

(3) Hedonism is false.

In a nutshell: Hedonism entails something false, so hedonism is false.

How does Nozick argue for (1)? Some philosophers have suggested that he argues for it by appeal to a claim about what we would want or choose to do if we were given the option of plugging in, in the following sort of way:

We would not want or choose to plug in to the machine, and this makes it the case that plugging in would not be best for us.8

Some who attribute this argument to Nozick object that nothing follows from the fact that something is desired (or would be desired under certain conditions) when it comes to whether it is desirable (i.e., worthy of being desired).9

Others point out that if Nozick’s objection to hedonism includes this argument, then it begs the question against hedonism by presupposing that well-being is determined by something other than pleasure and pain – namely, desire satisfaction and frustration.10

These worries miss the mark, however, since Nozick never intended to argue for (1) in this way. He explicitly disavows this argument here:

Notice that I am not saying simply that since we desire connection to actuality the experience machine is defective because it does not give us whatever we desire…for that would make “getting whatever you desire” the primary standard. Rather, I am saying that the connection to actuality is important whether or not we desire it—that is why we desire it—and the experience machine is inadequate because it doesn’t give us that.11

Not only, then, is Nozick not appealing to a desire-based theory of well-being in his objection to hedonism, he intends the machine to make trouble for desire-based theories as well (a point I will return to in Section 4).

Why, then, does Nozick ask us to consider what we would want or choose to do at all? The most charitable answer is: merely as an intuition pump for (1). That is, he asks us to consider whether we would want to plug in as a way of getting us to have the intuition that plugging in would not be best for someone. Imagining oneself faced with the choice of whether to plug in, and seeing what one would want or choose to do in this scenario, makes vivid the fact that it would not be in the best interests of a normal human being to plug in.

What, then, is Nozick’s argument for (1)? It may be suggested that he argues for (1) by pointing out some of the things that a person would be missing out on by plugging in – for example, in Nozick’s words, the ability ‘to do certain things, and not just have the experience of doing them’,12 to ‘focus on external reality, with [one’s] beliefs, evaluations, and emotions’,13 to explore ‘reality and [respond], altering it and creating new actuality ourselves’,14 and so on.
But this suggestion, too, seems to mistake Nozick’s intention. While Nozick does indeed say that it is for reasons such as these that plugging in would not be best for one, this is not part of some *argument* that he has for (1). Instead, he seems to think that reflection on the experience machine case yields *two* distinct revelations – on the one hand, that hedonism and desire-based theories are false, and on the other, that well-being includes something like an ability to connect with or interact with reality.

Nozick, I believe, does not attempt to argue for (1). Instead, he takes it for granted that most of his readers will find (1) intuitive. This has proven to be a safe assumption. Even those sympathetic to hedonism have admitted to finding (1) intuitive. The genius of Nozick’s argument lies simply in pointing out that something interesting and contested (i.e., the falsity of hedonism) appears to follow from something that is found almost universally acceptable (i.e., (1)).

Before moving on, it is worth noting that, while the objection I have attributed to Nozick here is the important one that has loomed so large in recent literature on well-being, not everyone is convinced that it is Nozick’s own. The chief dissenter is Feldman (2011). Feldman considers roughly the interpretation of Nozick I have offered and says of it: ‘Possibly an interesting argument; definitely not in the text.’ His reasoning is as follows:

Careful study of the passage will reveal that Nozick does not explicitly claim to be refuting any theory of welfare or of value in general. He never mentions welfare or wellbeing or value or intrinsic value in the passage. Instead, he speaks almost exclusively about certain psychological matters. Thus, for example, he says (p. 43, 44) that reflection on the Experience Machine teaches us something about “what matters to us” or what is “important to us”. In other places he suggests that it tells us something about what we desire (p. 43), or what we would choose. All of these remarks more strongly hint that he was interested in a psychological claim about what we value rather than in an axiological claim about what is valuable.

But Feldman is here basing his interpretation of Nozick’s objection solely on the text of *Anarchy, State, and Utopia*. Oddly, he admits that Nozick’s ‘remarks in later writings tend to suggest’ the interpretation he (Feldman) finds so implausible. He notes, for example, that in *The Examined Life*,

[Nozick] explicitly says that the example of the Experience Machine is intended to shed light on a question about value. In this context [Nozick] mentions the idea that “plugging in constitutes the very best life”.

Feldman also concedes that ‘it is possible that when Nozick says that something ‘matters to us’ he means not just that we care about it, but that it is in fact good for us.’

Finally, Feldman admits that the objection I have attributed to Nozick is ‘fairly interesting’, while the alternative interpretations of Nozick he considers are pretty clearly ‘bad arguments’.

In light of these points, not to mention my earlier observation that we can interpret Nozick’s appeal to what we would want or choose to do if given the option of plugging in merely as an intuition pump for (1), it seems most charitable to ascribe to Nozick the objection as I have outlined it here.

3. Recent Criticisms of Nozick’s Objection

In this section, I will explain and briefly discuss the most important recent criticisms of Nozick’s objection.
3.1. A REASONABLE FEAR OF CATASTROPHE

A common criticism of Nozick’s objection is that, while it is true that most of us would not want or choose to plug in to the experience machine, the most straightforward explanation of this fact is one that is entirely consistent with hedonism – namely, that we would fear, quite reasonably, that the machine might malfunction, that the premises on which it is kept might be overrun by fundamentalist zealots, that the scientists running the machine might turn evil, etc. If any of these things were to happen, then plugging in would be very bad for us indeed, whatever theory of well-being were true (including hedonism).22

However, as I claimed above, Nozick’s appeal to what we would want or choose to do if given the option of plugging in seems to function merely as an intuition pump for (1). It is a dispensable part of the objection. We can ignore the question of whether we would want or choose to plug in, and instead ask directly whether it seems best for someone – either oneself, a loved one, or a complete stranger – to plug in (or to have plugged in) in a case where the machine, as a matter of fact, does not malfunction, the premises are not overrun by fundamentalist zealots, the scientists do not turn evil, etc. Most people who ask themselves this latter question find themselves answering ‘no’ to it.23

3.2. AN IRRATIONAL FEAR, REVULSION, OR BIAS

A second criticism of Nozick’s objection is that our unwillingness to plug in might be due, not to a reasonable fear of catastrophe, but to an irrational fear, revulsion, or bias. Perhaps we’d refuse to plug in because we’d be too scared of having wires inserted into our skull (just as we might be scared of a spider we know to be venomless and so harmless), or because we’d ‘feel icky’ at the thought of being submerged in a vat of liquid, or because – as quite a few philosophers have suggested recently – we have a status quo bias, an irrational tendency to prefer the way things are now to new or different ways.24

As evidence for this latter (‘status quo bias’) explanation, it has been suggested that we consider a variant on Nozick’s case in which we are told that we are already plugged in to an experience machine. Ask yourself: Would you want to unplug? Many people say ‘no’.25 This is supposed to show that in Nozick’s original case, it might not be our picking up on the intrinsic value for us of contact with reality that explains our reluctance to plug in, but merely our having an irrational impulse to stick with what we already have or know.

These criticisms, however, also seem to mistake the function of Nozick’s appeal to what we would want or choose to do if offered the chance to plug in. Nozick could gladly accept that an important part of the reason we would be unwilling to plug in is that we have an irrational fear, revulsion, or bias – that we find the thought of plugging in too scary, icky, or alien. His gripe with hedonism stands: it does not seem best for someone to plug in to the machine.

3.3. THE DIFFICULTY OF HAVING A FINE-GRAINED INTUITION

A third criticism of Nozick’s objection is that it is very hard, or perhaps even impossible, when imagining the case, to have an intuition that is strictly about well-being (i.e., about whether it would be best for someone to plug in).26

If someone were to plug in, then this would likely be very bad for this person’s friends and loved ones, since they would be deprived of her company. Consequently, it may be morally wrong for this person to plug in, or best simpliciter that she not do so – things that might entail that she has most reason not to plug in (whether or not it would be best for her to plug in). Perhaps we are implicitly aware of these things when imagining the case, and our awareness of them makes it hard or impossible for us to have an intuition that is strictly about well-being. While
we may think that it is (1) we are finding intuitive, it might instead be something else—say, that a normal human adult would have most reason not to plug in.

Nozick himself was aware of this worry. This is why he suggested that in imagining the case, it might help to imagine also that other people

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.).

While the matter is hardly resolved, most commentators seem to share Nozick’s view that we can have an intuition that is strictly about well-being.

3.4. DEBUNKING THE INTUITION

A fourth criticism of Nozick’s objection is that, while (1) is indeed intuitive, this intuition is itself best explained, not as a response to or reflection of reality, but in a way that makes clear that it is not truth-tracking. When we understand where this intuition comes from, we will see that it should be ignored.

One possibility, for example, is that this intuition is the product of an irrational fear, revulsion, or bias. (Note the difference between this possibility and the earlier-mentioned possibility that our wanting or choosing not to plug in might be due to an irrational fear, revulsion, or bias.) Why do we feel that it would not be best for someone to plug in? Perhaps it is because we find the prospect of plugging in scary, repulsive, or alien, and our emotional response here is clouding or contaminating our intuition concerning well-being.

Another possibility is that our well-being intuition is being contaminated, not by an irrational fear, revulsion, or bias, but by our intuitions or judgments concerning other matters, such as the intuition or judgment that it would be morally wrong for one to plug in, that it would not be best simpliciter that one plug in, that one would have most reason not to plug in, that a life plugged in to the machine would be meaningless, aesthetically poor, low in achievement value, etc.

A further possibility is suggested by Matthew Silverstein. According to Silverstein, ‘[o]ur [well-being] intuitions tend to reflect our desires and preferences’, and we have been conditioned to have a strong ‘desire to remain connected to the real world, to track reality’. Why have we been conditioned to have this desire? Roughly, it is because having this desire tends to lead one to happiness, and desires that tend to lead one to happiness thereby get stronger.

At the same time, desires that tend to lead one away from happiness thereby get weaker, and the desire for happiness itself is among these latter desires. Over time, then, our desire for happiness itself gets weaker, and along with it (since our intuitions reflect our desires) our pro-hedonism intuitions.

Roger Crisp suggests a further explanation of the intuition that (1) is true. He says that wanting or caring about authenticity in relationships would likely have boosted the reproductive fitness of early humans. He writes:

Valuing honesty, transparency, genuineness, and so on, has a clear pay-off: it fends off deception, and thereby assists understanding of the world, which itself issues in a clear evolutionary advantage.

Moreover, feeling or judging that accomplishment mattered in itself would likely have made one a better hunter during the Stone Age, and so resulted in one’s being ‘rewarded by [one’s] fellows, partly with a larger share of the available goods, but also with esteem and status within the group’.
Perhaps one reason we find (1) intuitive is that we are descended from those early humans who survived and reproduced so effectively, in part, by happening to have these sort of non-hedonistic dispositions and beliefs.

These criticisms of Nozick’s objection merit serious consideration. In the remainder of this section, I want to raise three important challenges for them. First, it would be helpful if these authors were to explain exactly what the process is by which our well-being intuitions are affected by the desires, values, dispositions, or beliefs in question. Suppose Silverstein is right that we are likely to arrive at a powerful desire for contact with reality via the processes he describes. How does this desire in turn lead to our having a pre-theoretical feeling that contact with reality is intrinsically good, not only for oneself, but for people more generally? Or suppose Crisp is right that valuing authenticity, or wanting to really achieve things (as opposed merely to enjoying the fruit of one’s achievements), tended to boost the fitness of early humans. How does this valuing or desiring lead one in turn to feel (again, at a pre-theoretical level) that authenticity and accomplishment are good for those who possess them? What is the process?

Second, it would be helpful if these authors were to explain what intuitions are not contaminated in the ways they describe. Presumably, we need to rely on some well-being intuitions in theorizing about well-being. Which ones? If Silverstein is correct that our well-being intuitions tend to come from our desires or preferences, are there any such intuitions that do not? If so, which ones? Alternatively, are there any basic desires that are themselves responses to or reflections of reality? Silverstein seems to think that the desire for pleasure or happiness may be one such desire, but his argument for this claim is underdeveloped—it would be nice to hear more on this.34 If Crisp is right that we can ignore well-being intuitions that we have only because it was fitness-enhancing for our ancestors to have certain dispositions or beliefs, are there any intuitions that do not fall into this category? Which ones?

Third, there seem be some people who find (1) intuitive but who do not have any of the desires, values, beliefs, and dispositions these authors cite. I, for example, find (1) highly intuitive, but I doubt very much whether I intrinsically want contact with reality, or value authenticity and accomplishment in the ways Crisp mentions. How, then, did I, and others like me, come to find (1) intuitive?

3.5. DISPUTING (2)

While most critics of Nozick’s argument have focused on the status of the intuition that (1) is true, there is a different route available: dispute (2). Hedonism might not entail that plugging in would be best for someone. Hedonism might, in other words, be consistent with its not being best for one to plug in.

One hedonist, Feldman, has suggested a version of hedonism that would deliver this result. On Feldman’s truth-adjusted hedonism, the amount that a given pleasure adds to its subject’s well-being depends on whether it is taken in something true.35 Pleasures taken in true things add more to well-being than otherwise identical pleasures taken in false things. Most pleasures that are experienced in Nozick’s machine would be taken in false things – one would think one was succeeding in one’s career, living with the most amazing partner, traveling the world, and so on, when none of this would really be happening. Therefore, according to truth-adjusted hedonism, these pleasures might add less to well-being than similar pleasures that were experienced in the real world.

A second possibility is that (2) is false because one could not survive plugging in. Plugging in, after all, would not give one any pleasures if it resulted in one’s death. Nozick himself says that plugging in constitutes ‘a kind of suicide’, and of somebody plugged in to the machine, ‘there is no way he is’.36
Why might plugging in kill one? Perhaps it is because, if one agrees to plug in, the machine would have to erase one’s memory of choosing to plug in, and this form of mental tampering might interrupt one’s psychological continuity in such a way that one’s consciousness would come to an end and be replaced with a numerically distinct one.

But this suggestion, quite apart from its radical claim that the form of mental tampering in question would literally kill one, seems unable to explain why being plugged in by somebody else, without one’s knowledge (say, while asleep) would still seem not best for one. If I am plugged in without my knowledge, there is no need for the machine to tamper with any of my memories.

A third possibility is that, while plugging in might not kill one, there are certain pleasures that no machine like Nozick’s could give one. For example, the pleasures of autonomy or free action may require the actual exercise of free will, something that is impossible in the machine (perhaps the machine can give one at best the impression of acting freely – a pale imitation of the real thing).

But presumably, the machine could be set up in such a way that it works, not by merely playing one a video tape of a life, as it were – including the false appearance not only of having various options, but also of choosing freely among them – but by improving one’s apparent options (i.e., one’s options as they appear to one). If the machine were set up in this latter way, it would still be required that one choose among various options. So, one would still be capable of exercising a kind of free will, and so (even if the pleasures of free action require the actual exercise of free will) have access to the associated pleasures.

Another suggestion is that the machine could not give one the full range of the pleasures of love and friendship. People often say that an important reason they would not plug in to the machine is that it would involve permanent separation from their friends and loved ones. Nozick himself writes:

…we want a connection to actuality that we also share with other people. One of the distressing things about the experience machine, as described, is that you are alone in your particular illusion. (Is it more distressing that the others do not share your “world” or that you are cut off from the one they do share?)

Perhaps the reason permanent separation from one’s friends and loved ones would be so bad for one is that it would necessarily have experiential consequences for one. The pleasures of love and friendship may require a certain subtlety in the language, facial expressions, bodily gestures, and actions of those around one that is beyond the capability of AIs (or at least AIs that fall short of real conscious selves – the sort that would populate Nozick’s machine).

A general problem for this third account of why hedonism is consistent with (1) is that, whatever pleasures one would be unable to get in the machine (whether of free action, love and friendship, etc.), these pleasures would have to be so very valuable for one that their absence could not possibly be compensated for by the very many pleasures that one surely could get in the machine.

4. The Desire-based Theorist’s Explanation

Many who claim that Nozick’s machine refutes hedonism accept some form of desire-based theory of well-being – i.e., the theory on which lives can be good or bad for their subjects just in virtue of their ability to get (and fail to get) what they want. According to these philosophers, while hedonism cannot account for why plugging in would not be best for one, desire-based theories can. This is because most of us want contact with reality – or, at least, real accomplishment, real friendship, etc. – and plugging in would frustrate these desires.

As I noted earlier, Nozick himself explicitly considered and rejected this account of why plugging in would not be best for one. According to Nozick, the reason we would want
not to plug in is that we would realise (even if only implicitly) that plugging in would not be best for us. The reason that plugging in would not be best for us is not that it would deprive us of things that we want, but that it would deprive us of things that we should want even if we do not.

Nozick’s opinion aside, it is worth questioning the adequacy of the desire-based theorist’s explanation. One reason for thinking it inadequate is that seemingly not everyone has an intrinsic desire for contact with reality (or for things such as real accomplishment, real friendship, etc.), yet intuitively, even those who lack such desires would still be missing out on something by plugging in. When we encounter those rare individuals who say they would not mind plugging in, or would even welcome the opportunity to do so, we tend not to feel ‘Oh well, plugging in would be best for them.’ Instead, we tend to feel that these people are making some kind of mistake – and not simply because they do not properly understand their own preferences (whether actual or idealized).39

In light of both Nozick’s opinion of the matter and the serious worry I have just mentioned for the desire-based theorist’s explanation, it may be more accurate for philosophers to start thinking of Nozick’s experience machine as an objection, not to hedonism in particular, but to hedonism and desire-based theories taken collectively.

5. Conclusion

In this article, I have had three main goals: First, to reconstruct Nozick’s objection. Second, to explain and briefly discuss the most important recent criticisms that have been made of it. Third, to question the conventional wisdom that the experience machine, while it neatly disposes of hedonism, poses no problem for desire-based theories of well-being.

Acknowledgements

I wish to thank an anonymous referee and the editor at Philosophy Compass for their invaluable feedback on this paper.

Short Biography

Ben Bramble obtained his PhD in philosophy from the University of Sydney in 2014. He is currently a Postdoctoral Fellow at Lund University, Sweden. His main research interests are moral philosophy, applied ethics, and political philosophy.

Notes

* Correspondence: Lund University, Sweden. Email: b.e.bramble@gmail.com

1 For a useful discussion of the concept of well-being, see Campbell (2015).
2 While rough, this definition will suffice for present purposes.
4 Contemporary advocates of hedonism include Feldman (2004), Crisp (2006), Heathwood (2006), Bradley (2009), and Bramble (forthcoming).
5 For an extensive list of authors who have ‘stated or implied that the experience machine thought experiment is a knock-down refutation’ of hedonism, see Weijers and Schouten (2013).
8 See, for example, Kawall (1999), Baber (2008), Silverstein (2000), Hewitt (2010).

See Baber (2008).


See, for example, Bradley (2009), p. 10.


Crisp (2006), Hawkins (2015), and Lin (forthcoming A) suggest that Nozick could deal with this first criticism by dropping the appeal to a choice situation altogether, and instead having us consult our intuitions concerning the respective levels of well-being of two individuals whose lives have been experientially identical from birth to death, but where only one is connected to reality. But this amendment seems to me to substantially weaken Nozick’s objection, as it seems much more intuitive that these experientially identical lives are equal in well-being than that it would be best for someone halfway through her existing life to plug in.

See, for example, Kolber (1994), De Brigard (2010), and Weijers (2014).

See De Brigard (2010).

See Hewitt (2010).


See, for example, Sobel (2002), Glover (1984).

See, for example, Heathwood (2006), p. 553.

Silverstein (2000).


For a similar criticism of Silverstein, see Lin (forthcoming A).


I develop this suggestion further in Bramble (forthcoming)

For a related point, see Lin (forthcoming B).

Works Cited


Bramble, B. ‘A New Defence of Hedonism about Well-Being,’ *Ego* (forthcoming)


Lin, E. ‘How to Use the Experience Machine.’ Utilitas (Forthcoming A).
—— ‘Against Welfare Subjectivism.’ Noûs (Forthcoming B).