What We Learn from the Experience Machine*

Fred Feldman
University of Massachusetts at Amherst

1. A Little Passage with a Big Impact

Nozick discusses the Experience Machine in a little (four paragraph) section of Chapter 3 of Anarchy, State and Utopia.¹ The section seems to be a mere speculative digression from Nozick’s main line of argument, and yet it has come to be perhaps the most widely discussed passage of the book. It has been reprinted and paraphrased and discussed thousands of times. Yet it seems to me that the passage is a bit of a mystery – perhaps it functions as a kind of Rorschach test for readers. Different readers apparently find different arguments in Nozick’s remarks. It may seem that the various interpretations tell us more about the interests of the readers than about the argument that Nozick actually presented.

In Section 2 I outline the context in which the controversial passage appears. What I say in this part may come as a surprise to anyone whose knowledge of the passage derives entirely from seeing it only as a selection in an anthology, isolated from its original context. After sketching the contents of the passage in Section 3, I go on in Section 4 to explain and evaluate some versions of one fairly popular interpretation of the argument. Under these interpretations, the argument is taken to be an attack on utilitarianism. I claim that every one of these interpretations is implausible. According to a much more popular style of interpretation, Nozick’s argument was intended to refute ethical

---

¹I am very grateful to Kristian Olsen for generous assistance on earlier drafts. I am also grateful to Ralf Bader, Owen McLeod, and Chris Heathwood for extensive critical comments and suggestions.

ⁱ He returns to the topic later in The Examined Life, Philosophical Explanations, “The Pursuit of Happiness” as well as in other writings. I will focus here on the passage in AS&U.
hedonism (or perhaps the entire class of “mental state” theories of welfare). In Section 5 I present several versions of the argument so understood. I go on to explain why I think every such version of the argument fails. A few commentators have supposed that Nozick was attacking some form of psychological hedonism. I discuss their views in Section 6. Finally, in Section 7, I summarize my views on all the arguments that are discussed in the paper. In a short appendix I offer – briefly and tentatively – a somewhat different argument that is merely suggested by Nozick’s remarks.

2. The Context

Nozick’s project in the first sections of AS&U is straightforwardly at the heart of political philosophy: he’s trying to show how the existence of a state could be justified. He’s trying to do this by starting with a description of a certain anarchic situation. In this situation, people have certain inviolable moral rights – fundamentally the right not to be attacked or killed and the right not to have their possessions stolen. He imagines that voluntary “protective associations” might spring up in this state of nature. People might get together for mutual protection, each one pledging to the others that if they are attacked, he will come to their aid. This might be a good idea. But such protective agencies would not be states; they would just be voluntary associations of individuals who have banded together for one fundamental purpose: mutual protection.

It’s possible that a “dominant protective agency” would arise in some area. This one would be the most powerful in its geographical region. It would have the power to protect its members (for a fee) and could insist on various concessions from members. But even such a thing would not be a state. One crucial factor is voluntariness. Suppose some individual is residing in the domain of a protective agency. Suppose he chooses to fend for himself, and thus not to make use of the agency’s services. Then the agency cannot force him to join. It’s voluntary after all. Furthermore, it is under no obligation to protect the interests of non-members living within its domain. Such an agency could not claim any monopoly on the determination of who can use force and when they can use it. The agency would have no power to say, for example, that other groups living within its
domain cannot form independent militias. A state, on the other hand, apparently would be able to do these things.

The Nightwatchman State (or Minimal State) goes well beyond this sort of dominant protective agency. It does offer protection to everyone in its domain. If some can’t pay for this protection, it provides tax-funded vouchers. Thus, it takes money from some members so as to pay the costs involved in defending others. It thus appears to be redistributive.² It compels some people to pay for benefits to be enjoyed by others. This is a state, though of course a minimal one.

Nozick claims that there is a form of state-like arrangement that is intermediate between the dominant protective agency and the minimal state. He calls this the “ultramiminal state”. Unlike the mere protective agency, the ultraminimal state does claim a monopoly on the use of force. It prohibits the formation of other protective agencies (such as the Mafia, or the KKK) within its domain. However, unlike a full-blown Nightwatchman State, it does not provide services to those who don’t join or can’t pay. It is still a “private” association. Nor does it provide any other services to members. It takes money from members and provides protection (but nothing else) to members.

Nozick wonders whether an advocate of the ultraminimal state might be open to charges of inconsistency. What could justify minimal redistribution for protective services but not for anything else? The advocate of such a state seems to think that there is something especially important about the protection of members’ rights not to be harmed; but at the same time he seems to think it’s OK to violate these rights when exacting payments from members to set up a system to protect the rights of others. So he advocates the violation of the rights of some in order to protect certain of the rights of others. Is this coherent?

This is the context in which Nozick starts talking about the utilitarianism of rights. He sees two options: the non-violation of rights might be taken as a goal. In this case, it would be permissible to violate some rights if that were required in order to ensure that

² Although, as Ralf Bader reminded me, Nozick elsewhere claims that this appearance might be misleading.
rights as a whole would be minimally violated. On the other hand, the non-violation of rights might be taken as a constraint. In this case, the violation of someone’s rights would be forbidden even if as a result many others would have their rights violated.

If a proponent of the ultraminimal state takes non-violation of rights to be the goal, with violations of such rights to be minimized as the utilitarianism of rights would suggest, he is in a bind. For in this case he is taking the non-violation of rights as a goal; and he is willing to adopt any means necessary to achieve this goal – even if it involves violating someone’s rights. This may seem to be an incoherent conception. But on the other hand, if he accepts a side constraints view, there is no inconsistency. For in this case he does not advocate violating anyone’s rights for the purpose of defending the rights of others.

This leads Nozick to a discussion of the question whether the side constraints view should be understood to include a constraint against violating the rights of animals as well as those of humans. And this leads him into a somewhat unexpected discussion of vegetarianism. And this leads him into an even more unexpected discussion of something like the non-identity problem (later discussed more directly by Parfit). And all this leads him to a consideration of a surprising moral principle: “utilitarianism for animals; Kantianism for people”. Nozick sketches this principle as follows:

UA/KP: (1) Maximize the total happiness of all living beings; (2) place stringent side constraints on what one may do to human beings. (p. 39)

UA/KP does not include a constraint against violating the rights of animals. It would be permissible to violate their rights if that were required in order to maximize happiness overall.

Nozick mentions “a thicket of questions” about utilitarianism for animals. Some of these are reminiscent of questions that were subsequently made famous by Parfit (though of course Parfit did not restrict himself to consideration of non-human animals). For example, does the theory imply that there is a duty to increase the number of animals
even when each animal is less happy just so long as the total amount of happiness is increased? (The Repugnant Conclusion) Would it be morally permissible for someone to kill off everyone else if that made him ecstatic – after all, the average level of happiness would thereby be maximized? (The Ten Little Indians case or perhaps the Utility Monster case.) Is it OK to kill one person provided that you immediately replace him with another person who is slightly happier? (The Non-Identity Problem).

Those are all tough nuts for the utilitarian to crack. They raise serious questions about utilitarianism for people. However, it remains possible that utilitarianism for animals would be acceptable. And this at last leads to the discussion of the Experience Machine. If there were such a machine, would we choose to plug in? It seems to be a digression within a digression within a digression. Nozick strongly suggests (p. 45) that he is discussing it for one main reason: we need to know what matters for people in addition to their mere experiences, and we need to know whether this thing (whatever it may turn out to be) also matters for animals, and so long as we have not answered these questions, we “… cannot reasonably claim that only the felt experiences of animals limit what we may do to them.” (p. 45)

There follows the famous passage.

3. The Passage Itself

The section entitled “The Experience Machine” consists of only four paragraphs. The thought experiment described in the first three paragraphs is nowadays familiar to virtually all philosophers. I will provide only a reminder. The reader is encouraged to reread the passage itself.

In the first paragraph (pp. 42-3) Nozick briefly describes the machine. He says it would “give you any experience you desired”. He mentions the experiences of writing a great novel, making a friend, or reading an interesting book. “All the time you would be
floating in a tank, with electrodes attached to your brain.’’ He then asks “Should you plug into this machine for life…?” (p. 42, emphasis added.) He then describes how you could emerge from the tank after two years and pick a new set of experiences that you could enjoy upon your return to the tank. Then he asks “Would you plug in?” (p. 43, emphasis added.)

In the second paragraph (p. 43) Nozick mentions two things that matter to us in addition to our experiences: “First, we want to do certain things, and not just have the experience of doing them.” Second, “… we want to be a certain way, to be a certain sort of person.” He claims that if a person plugs into the Experience Machine, he commits “a kind of suicide.” “… there is no way he is.”

In the third paragraph (pp. 43-4) Nozick says that if you plug into the machine, you “would be limited to a man-made reality, to a world no deeper or more important than that which people can construct.” He proceeds to sketch the analogy between plugging into the machine and taking psychoactive drugs. In each case, Nozick acknowledges that readers may reach different assessments: does the subject merely “surrender”, or does he take an “avenue to a deeper reality”?

In the fourth paragraph of the section, Nozick engages in speculation about a “transformation machine” and a “result machine”. Nozick says that he will ‘not pursue here the fascinating details of these or other machines.’ He goes on to offer some speculations:

“Perhaps what we desire is to live (an active verb) ourselves, in contact with reality. (And this, machines cannot do for us.) Without elaborating on the implications of this,

---

3 It’s interesting to compare Nozick’s description of the subject in the Experience Machine with J.J.C. Smart’s discussion of the “bald-headed man with electrodes protruding from his skull”. In both cases, the electrodes are imagined to be connected to a machine in such a way as to induce pleasant experiences. And in both cases, the question is whether the life of the subject would be a good one, or one we would desire. Smart’s discussion of the “electrode man” first appeared in his 1961 book An Outline of a System of Utilitarian Ethics. The passage reappears on p. 19 of his contribution to Utilitarianism: For and Against, which was published in 1973. AS&U was published in 1974. Nozick does not mention Smart in Section 3 of AS&U, and he does not cite Smart in his bibliography or index.
which I believe connect surprisingly with issues about free will and causal accounts of knowledge, we need merely note the intricacy of the question of what matters for people other than their experiences. Until one finds a satisfactory answer, and determines that this answer does not also apply to animals, one cannot reasonably claim that only the felt experiences of animals limit what we may do to them.” (AS&U, pp. 44-5)

Different readers come away from this section with different ideas about what Nozick was trying to show. Most philosophers who comment on the passage seem to agree that Nozick presented the example of the Experience Machine in order to show that some philosophical or psychological theory is false. There is disagreement about the target theory: precisely what was Nozick out to refute? I think there are four main interpretations. According to some, his target was (A) act utilitarianism. According to others it was (B) ethical hedonism as a theory of individual welfare. Others seem to think he was attacking (C) the whole class of mental state theories of welfare. Still others think the target is not in ethics at all. Rather, it is (D) psychological hedonism – a theory about human motivation.

Let’s first consider how the argument might be thought to constitute an attack on act utilitarianism.

4. Understood as an Argument Against Utilitarianism

---

4 In addition to the four views mentioned here, there are some “outliers”. Sober and Wilson seem to claim that Nozick appealed to the Experience Machine in an effort to refute psychological egoism. See Sober and Wilson (2000): 199. Another commentator says that Nozick was trying to show that a person’s welfare level is not determined by how happy he is. He takes the target to be eudaimonism. See Jollimore (2004): 333. Griffin says (1988): 9 that the target is a sort of hybrid of hedonism and preferentism as a theory of welfare.

5 I should mention as well that some commentators claim that the example of the Experience Machine is designed to establish a positive thesis about welfare. “This thought experiment [the experience machine] shows that a valuable life involves certain character traits, the exercise of certain capacities, and having certain relations with others and to the world and, hence, that value cannot consist in psychological states alone.” David Brink, Moral Realism and the Foundations of Ethics, p. 224.
The discussion of the Experience Machine comes right after some talk about “utilitarianism for animals”; immediately following it there comes a section in which Nozick mentions that utilitarianism allows us to treat animals as mere means. Perhaps these facts about context have led some commentators to think that Nozick was presenting an argument against utilitarianism in the Experience Machine passage. There is no evidence in the passage itself that would support this interpretation. In light of things he says elsewhere in the book, it seems clear that Nozick thought that utilitarianism was in deep trouble. He mentions a series of familiar objections. None of these has anything to do with the Experience Machine. But since so many apparently think that Nozick was attacking utilitarianism, and since he does at least talk about it in the vicinity, it may be worthwhile to consider whether there are hints of an interesting argument against utilitarianism in the Experience Machine passage.

When we speak of utilitarianism, we are really speaking about a whole family of similar views. Each of these views is a general theory about morally right action. Each purports to state necessary and sufficient conditions for the moral rightness of specific act tokens. Different forms of utilitarianism can be seen as variations on this central theme:

\[ \text{AU: An act token, } a, \text{ is morally right if and only if } a \text{ maximizes actual total utility.} \]

In one place near the beginning of the passage (p. 42) Nozick rhetorically asks “Should you plug into this machine for life, preprogramming your life’s experiences?” The context makes it clear that he thinks the obvious answer is ‘No. You should not plug in.’ This gives us a hint of one way to understand the argument.

---

6 This sort of interpretation seems to be endorsed by MacNiven (1993): 4; Railton (1984): 148-9; Teichman and Evans (1999): 91-2; among others. Smart’s example of the electrode man was explicitly intended to figure in an objection to utilitarianism; perhaps some readers assume that Nozick must have been following suit.

7 In a brief passage (41-2) Nozick alludes to the repugnant conclusion, the Ten Little Indians, person-affecting problems, problems about the evil of killing. Elsewhere he makes clear his objections to utilitarianism as a theory of distributive justice. None of these objections turns essentially on intuitions brought out in the Experience Machine passage.

8 One variant would make use of the concept of satisficing instead of maximizing; another variant would employ the notion of expected utility instead of actual utility. There are dozens of variations.
Let’s say that you are in the “Nozick Scenario” if you have been living a fairly pleasant life and can reasonably expect to go on in pretty much the same way; but someone has offered you the opportunity to plug into an Experience Machine. You have been assured that if you plug in, you will experience a much greater balance of pleasure over pain, just as Nozick has described. Now we can state the argument:

**First Anti-Utilitarian Argument**

1. If AU is true, then you should plug into the Experience Machine when in the Nozick Scenario.
2. But in fact it’s not the case that you should plug into the Experience Machine when in the Nozick Scenario.
3. Therefore, AU is not true.

This argument is unworthy of extended discussion. There is nothing to be said for premise (1). It seems unlikely that a person floating his life away in a tank would be doing much to enhance the utility levels of his former friends and family. Although his own utility level might be fairly high, he would be missing out on many opportunities to provide utility for others. Hence, in anything approaching normal circumstances, the act of plugging in would have relatively low utility. AU then implies that you should not plug in. Premise (1) is just silly.

In the rest of the Experience Machine passage, Nozick does not say that you should not plug in; rather, he says (or strongly implies) that you would not plug in. This makes it considerably more difficult to see how the premise could be thought to bear on AU. Surely premise (1) of this argument is too preposterous for any discussion:

**Second Anti-Utilitarian Argument**

---

9 Nozick comments on this point, first in AS&U and then later in The Examined Life. I discuss the relevance of his remarks below in fn. 17.

10 I leave it to the interested reader to determine whether any other form of act utilitarianism would be refuted by an argument of this form. So far as I can tell, no recognized form of utilitarianism implies that you should plug in. Premise 1 remains false under all reinterpretations.
1. If AU is true, then you would plug into the Experience Machine when in the Nozick Scenario.

2. But in fact it’s not the case that you would plug into the Experience Machine when in the Nozick Scenario.

3. Therefore, AU is not true.

AU has no implications concerning what you would do under any circumstances.\textsuperscript{11}

Nozick’s reflections on the Experience Machine might be thought to be connected to utilitarianism in a different way. Perhaps the argument goes like this: you would not plug into the Experience Machine when in the Nozick Scenario. This shows that you value something other than pleasure. And this shows that something other than pleasure has value – in other words, it shows that hedonism is false. But utilitarianism is just the combination of (a) the idea that we should perform the act that maximizes utility, and (b) the idea that the utility of an act is determined hedonistically. Therefore, since hedonism is false, utilitarianism is false. (We may call this the Third Anti-Utilitarianism Argument.)\textsuperscript{12}

That would be a tidy little argument, moving smoothly from what seems to be Nozick’s central claim (“you would not plug in”) to a very interesting conclusion (“utilitarianism is false”). However, there are several serious problems with this interpretation. Foremost among them is that there is no clear evidence in the text that Nozick intended to present this argument. While he does seem to commit himself to the first premise, and he seems to endorse something like the suggested conclusion, there is no hint in the relevant section of AS\&U of the intermediate premises.\textsuperscript{13}

The second problem with the interpretation is that the proposed argument would be relevant only to versions of utilitarianism that incorporate a hedonistic axiology. Other

\textsuperscript{11} Again, I leave it to the reader to consider whether some other form of act utilitarianism might give rise to a more plausible version of premise 1.
\textsuperscript{12} Railton seems to interpret the argument in this way. See Railton (1984): 148-9.
\textsuperscript{13} Nor, I should note, is there any mention of the conclusion in the Experience Machine passage itself.
versions would be unaffected. Consider, for example, the sort of utilitarianism that ranks outcomes by appeal to the extent to which preferences are satisfied. The current anti-utilitarian argument cannot be revised so as to cast doubt on that sort of utilitarianism.

Consider, for another example, the sort of utilitarianism that ranks outcomes by appeal to the extent to which people realize their distinctive human potential. The proposed argument would cut no ice with respect to this theory, either. The same could be said for a wide variety of preferentist, eudaimonist, perfectionist, and pluralist forms of utilitarianism. Thus, in the absence of solid proof that this was the argument Nozick intended, it would be more charitable to seek some other interpretation. Perhaps the argument is meant to refute only those forms of utilitarianism that make use of a hedonistic axiology.

But this leads us to a deeper point about the dialectic. On the latest proposed interpretation, the Experience Machine argument is intended to refute hedonistic utilitarianism. The argument proceeds by casting doubt on the axiological component of the normative theory. It is silent concerning the basic idea that right acts are ones that maximize some sort of value. It seems to me, in this case, that it would make more sense to view the argument simply as an attack on hedonism. If it refutes hedonism, then by extension it would refute any wider theory that contains hedonism as a part.

5. Understood as an Argument Against Ethical Hedonism

Quite a few commentators have thought that Nozick’s target in the Experience Machine passage was ethical hedonism.\textsuperscript{14} For purposes of discussion here, I take ethical hedonism to be a theory about personal welfare or well-being. To be more precise, I take it to be the view that a person’s welfare is directly proportional to the amount of sensory pleasure-minus-pain that he experiences.

One immediate implication of this form of hedonism is that if one possible life has a greater balance of pleasure-minus-pain than another, then the one must have greater welfare value than the other, regardless of other differences between them. This gives us a neat way to understand the point that many readers find in the Experience Machine passage.

Imagine that you have started out living a certain life, LA (your Actual Life). Things are going reasonably well for you and if you stick with LA you will end up with a fairly pleasant life. Now you are in the Nozick Scenario and someone has offered you a chance to plug into the Experience Machine. He tells you that if you plug in, you will have many intense and long-lasting episodes of pleasure. You will get to live life LEM (Life on the Experience Machine), which will give you a significantly greater total amount of pleasure-minus-pain. With this as background, we may state a version of the argument that some claim to find lurking in the Experience Machine passage:

**First Anti-Ethical Hedonism Argument**

1. If you were in the Nozick Scenario you would not plug in so as to get LEM. You would stick with LA.

2. If (1), then LEM does not have greater welfare value than LA.

3. If LEM does not have greater welfare value than LA, then ethical hedonism is false.

4. Therefore, ethical hedonism is false.

I think (1) is plausible. At least, it is plausible when construed as a claim about what *I* would do if *I* were in the Nozick Scenario. If *I* were given such a choice, *I* wouldn’t choose LEM. For *me*, the choice would not be difficult. *I* would prefer to stick with my real life rather than take on a new life on the Experience Machine. I assume that many others are like me in this respect.\(^\text{15}\) But I think there is no reason to accept (2). From the

\(^{15}\) But this premise has been challenged. Torbjorn Tannsjø (1998): 112 points out that lots of people take mind-altering drugs. These drugs can be seen as pharmacological experience machines. They cause the subject to have feelings of pleasure associated with dreamy, delusional, hallucinatory, unrealistic, drug-induced experiences. Felipe de Brigard claims to have given some questionnaires to his students. In one
fact that you would be reluctant to choose LEM over LA, nothing follows about the welfare values of the two lives. There are several possible explanations for this. One of the main problems here is epistemic: how can you be sure that the machine would work as advertised? Maybe it’s just a scam. Maybe the machine would malfunction. This would give you good reason to avoid plugging in. Refusing to plug in under these circumstances would have no implications concerning welfare values.16

Let us revise the example so as to avoid the epistemic difficulties. Let us assume that when you are given the choice of LEM or LA, you know beyond the shadow of any doubt that LEM would contain a substantially greater amount of pleasure-minus-pain. Under these circumstances, would you plug in? We may assume that Nozick would say that you would not.

But even if I were absolutely certain about what would happen to me if I were to plug in, I still might choose to stick with my old life rather than take on a new life in the Experience Machine. I might do this even though in the imagined circumstances I would know for sure that life on the machine would give me greater pleasure-minus-pain. One simple reason is this: I might think it would be morally wrong to plug into the machine. I have made a bunch of promises to my wife and my daughter and my students and my dog. I might recognize that if I plug in, I will be unable to keep those promises. Under these circumstances, I would probably think it would be morally wrong to choose the pleasant life on the machine and so I would not plug in. But my refusal to plug in would show nothing about the truth of hedonism. Premise (2) remains false.17

________

16 A number of commentators have focused on this problem with the argument. See, for example, Sumner, (1996): 95. Sumner asks, ‘What happens if there is a power failure?’ Crisp (2006): 635 alludes to the same point. He dismisses the question about whether people would choose to plug in, since their choices might be influenced by “differing attitudes to risk”. Goldsworthy (1992): 18, emphasizes this point, saying that a reasonable person might choose not to plug in because of fear of “catastrophic, unimaginably horrible consequences of malfunction or abuse.” Sober (2000) makes a similar point.

17 For a persuasive presentation of this objection, see Jason Kawall, ‘The Experience Machine and Mental State Theories of Welfare,’ The Journal of Value Enquiry 33 (1999): 381-87. Sumner makes a similar point, noting that “… welfare tracks only one dimension of the value of a life.” Sumner, 1996: 96. See also William Shaw Contemporary Ethics: Taking Account of Utilitarianism (Malden: Blackwell, 1999): 51. A series of tests, he asked them if they would choose life on the experience machine over their actual lives. Though the testing was all make-believe, and one may assume that the students all knew that the whole business was little more than a joke, apparently quite a few of them said they would opt for the machine.
Let us introduce another adjustment. Let us assume that the subject of the experiment is not only well informed; let us assume in addition that he is completely selfish. Let us assume that he is concerned only about himself. Ordinary selfishness may not be sufficient here; we need to go a bit further: we must assume that the subject is “welfare selfish”. He cares only about his own welfare. He does not care about his “moral welfare”. Might such a person still refuse to plug in? Would that show that hedonism is false? In other words, what shall we say about this argument?

**Second Anti-Ethical Hedonism Argument**

1. If you were in the Nozick Scenario, and you knew beyond doubt that LEM would be more pleasant than LA, and you were welfare selfish, still you would not plug in so as to get LEM. You would stick with LA.
2. If (1), then LEM does not have greater welfare value than LA.
3. If LEM does not have greater welfare value than LA, then ethical hedonism is false.
4. Therefore, ethical hedonism is false.

Even with these conditions built into the assumptions, it still seems to me that (2) is open to doubt. One further problem is that you might refuse to plug in because you find the idea of life in the Experience Machine icky and repugnant. It might just seem too weird and unnatural. In this respect you might be like a patient who refuses some novel

---

similar point is made in Haslett (1996): 41, as well as in Silverstein (2000): 290. Silverstein points out (291) that when he returned to this topic in *The Examined Life* Nozick emphasized some details of his description of the Nozick Scenario and thereby overcame this difficulty. Specifically, Nozick asked us to imagine that we are given assurances that others will also have the option of plugging into Experience Machines, and so we need not worry about failing to fulfill our moral obligations if we choose LEM. (As Ralf Bader reminded me, Nozick had already made this point in *AS&U*, p. 43.) However, it seems to me that the imagined clarification of the Nozick Scenario does not overcome the difficulty. A person might still worry about fulfilling his moral obligations even if he were convinced that others would be plugging in. For example, consider someone who solemnly promised his mother that he would rescue her if her Experience Machine should happen to malfunction. Believing that others would be in Experience Machines (or would have the option of plugging in) would not relieve him of his feeling of obligation to keep out of the machine so as to be available for rescue operations, should they be necessary. And even more obviously, suppose he knows that others have been given the opportunity to plug in but have chosen to remain unplugged. His feeling of obligation to them would be unaffected. See Nozick (1990): 105.
medical treatment because it just seems disgusting. He might recognize that undergoing the treatment would be for the best, but he might prefer to remain “naturally” sick.

We can deal with this problem as well. We can stipulate that the subject in the Nozick Scenario has to be fully rational.\textsuperscript{18} He cannot allow worries about “ickiness” or “unnaturalness” to deflect him from the course that will maximize his own long-term welfare.

In spite of all this, it still seems to me that premise (2) is open to doubt. You might know for certain that plugging in would be very pleasant, and you might be rational and welfare selfish, but you still might refuse to plug in. This could happen if don’t believe in hedonism. Suppose, for example, that you believe in some form of perfectionism. You think that a person’s welfare is determined by the extent to which he conforms to some human ideal. You think this human ideal involves moral, intellectual, aesthetic, athletic and other achievement. Since you are fully committed to this view about the Good Life, you are not impressed by the offer of a life on the Experience Machine. “Why would I want that miserable life?” you ask. “I would be much worse off if I were to plug in.”

The fact that you would not choose LEM in the imagined circumstances might be thought to show that you don’t believe that LEM has higher welfare value than LA, but it does not show that this belief is true.

In order to avoid this sort of difficulty, we need to stipulate that the person in the Nozick Scenario is not already committed to some anti-hedonistic theory of welfare. For so long as we allow that he might be so committed, his choice of LA over LEM might be due to his views about welfare rather than to the fact that LA is actually the better life.

But here we run into profound difficulties. If we say that the subject in the Nozick Scenario simply has no views about what makes for a good life, then it becomes doubtful

\textsuperscript{18} It would be difficult to give a fully satisfactory account of what is meant here by ‘rational’. I beg the reader’s indulgence.
that he would have any reason to choose to plug in for LEM or to stick with LA. He would presumably have no basis for thinking one life would be better than the other. If we say that the subject accepts hedonism as his theory of welfare, then surely the subject would choose LEM. In this case, premise (1) would be false. If we say that he must accept some non-hedonistic theory of welfare, then he might choose LA, but this would show nothing about the truth of hedonism. The subject’s choice would be attributable to his acceptance of a non-hedonistic theory of welfare.

There is another option. We can say that the subject in the Nozick Scenario is “axiologically insightful” – he prefers one life over another only if the one is actually better for him than the other. This rules out the possibility that the chooser might refuse to plug in because he is in the grip of some faulty axiology.

Now we may restate the argument:

**Third Anti-Ethical Hedonism Argument**

1. If you were in the Nozick Scenario and you knew beyond doubt that LEM would be more pleasant than LA, and you were completely welfare selfish, and you were rational, and you were axiologically insightful, still you would not plug in so as to get LEM. You would stick with LA.
2. If (1), then LEM does not have greater welfare value than LA.
3. If LEM does not have greater welfare value than LA, then ethical hedonism is false.
4. Therefore, ethical hedonism is false.

Now, it seems to me, we have reached a turning point in the dialectic. With all these extra conditions in place, premise (2) is beginning to seem more plausible. In fact, I find it difficult to imagine how a person with all the traits described here could fail to choose the welfare better life. He cares only about getting the life that will be best for him; he is completely rational; his preferences match the actual values of the possible lives. Every

---

19 Thanks to Owen McLeod for encouraging me to pursue this line of argument.
rationale for choosing the less-good life seems to have been ruled out. So if he were to choose LA, that would support the notion that LA is the better life; and that would refute hedonism.

But once we make these changes in the description of the choice scenario, the plausibility of premise (1) begins to fade. Surely it is now open to the hedonist to say that a chooser like the one we have been describing would choose life in the Experience Machine. The hedonist would say that since pleasure is the good, any welfare-selfish, knowledgeable, rational person whose preferences match the true values of things would undoubtedly choose the life of pleasure.

It is pointless to insist that you would not plug in. What you would do is irrelevant. We would need know what a perfectly rational, welfare selfish, knowledgeable etc. person would do. You are not like that. The idea of a life on the Experience Machine probably seems totally icky to you.\textsuperscript{20, 21}

\textbf{6. Understood as an Argument Against Psychological Hedonism}

\textsuperscript{20} In his paper on this topic, Felipe De Brigard describes an experiment in which his students were asked whether they would choose to plug in to an Experience Machine. Before announcing the choices they would make, the students were given make-believe accounts of what they would get if they chose to plug in. De Brigard claims that the results of his experiment are relevant to Nozick’s argument. But it should be clear that no amount of this sort of empirical data would bear on premise (1) of the latest argument. The students (unless De Brigard’s classes are very unusual indeed) were not “completely rational”, or “completely welfare selfish”, and they surely were not “completely convinced” that they would get the life they chose. They knew perfectly well that it was all just a bit of fantasy. So far as I can tell, there is no empirical experiment that would either confirm or disconfirm (1). It’s a philosophical claim about the connections between rationality, knowledge, welfare-selfishness, and axiology, on the one hand, and choice on the other.

\textsuperscript{21} Some have taken the target of Nozick’s attack to be the whole class of mental state theories of welfare. (e.g., Kagan seems to say this on pp. 35-6 of \textit{Normative Ethics}, and he explicitly mentions Nozick in a note associated with this passage.) Mental state theories of welfare include sensory hedonism, certain versions of the “happiness theory” (aka eudaimonism), and all other theories that identify the ultimate atoms of personal value as “inner states”. Such theories imply that if two lives are alike internally -- if everything “feels the same” to the people living the lives -- then those lives must be alike in welfare value. Since I have already spoken at length about the implications of the argument interpreted as an argument against ethical hedonism, and what I said about hedonism would carry over for other mental state theories of welfare, I will say no more than this here: There is nothing in the Experience Machine passage that casts any serious doubt on mental state theories of welfare.
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 well-being 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. We could take him at his word; we could assume that he is just trying to point out that people care about some things in addition to their experiences of pleasure.

It is, of course, perfectly obvious that people care about all sorts of things – what’s for lunch, what sort of car they will purchase, who will be elected president, etc. So there would have been no need for Nozick to bring in an extravagant hypothesis in order to show that experiences of pleasure are not the only things that people care about. We already know that people care about a lot of different things. In order to be worthy of attention, the doctrine in question would have to be a thesis about what people ultimately care about, or about what matters to people ultimately or intrinsically. So understood, the doctrine would imply that if anyone cares about something other than his own experiences of pleasure, he cares about it – in the long run – because he cares about pleasure and he thinks the thing would bring him some pleasure.

---

22 He uses the word ‘best’ in one instance. And in the following section he mentions utilitarianism, though it is not clear how this is supposed to relate to what he says in the discussion of the experience machine. It is also interesting to note that in his commentary on the Experience Machine in his later (1990): 105 The Examined Life he explicitly says that the example of the Experience Machine is intended to shed light on a question about value. In this context he mentions the idea that “plugging in constitutes the very best life”. Ralf Bader (in personal correspondence) has pointed out several other instances in which, in later writings, Nozick suggested that he took himself to be presenting an argument about the value of a life in the Experience Machine passage.

23 These seem to be matters of psychology, but, as Matthew Silverstein emphasizes, 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. See Silverstein, (2000): 286.
We must not forget about pain. We should add that the only thing that people ultimately care about *avoiding* is pain.

All this talk about “care” and “importance” and “mattering” is understood to be connected to motivation. When someone cares about something, or when it matters to him, it figures in his motivation. Psychological hedonism may be taken to imply that at least in their serious and considered behavior, people are motivated – ultimately – by their desire for pleasure and their aversion to pain. We might even go so far as to say that psychological hedonism implies that people are always ultimately motivated by a desire to maximize their own pleasure-minus-pain. This ultimate motivation may be hidden – even the agent himself may fail to recognize that he is, ultimately, doing things that he (perhaps at some deeper level) thinks will maximize his hedono-doloric level.

With these thoughts in mind, we may state one form of psychological hedonism in this way:

**PH:** Ultimately, each person is motivated solely by a desire to maximize his own hedono-doloric level.

So understood, psychological hedonism is a general thesis in human psychology. It may be what Bentham had in mind in that famous passage where he spoke of the “two sovereign masters, pain and pleasure” who alone “determine what we shall do”.

Some commentators have assumed that Nozick appealed to the Experience Machine in order to refute psychological hedonism rather than ethical hedonism.\(^{24}\) Nozick never states anything quite like PH; at best his remarks about “what we care about” and “what

---

\(^{24}\) Elliott Sober discusses the Experience Machine at some length in his article “Psychological Egoism” (2000). He seems to think the argument can be used against psychological hedonism. In two papers on this topic, John Lemos consistently indicates that he interprets the argument in this way. See Lemos (2002) and Lemos (2004). Felipe De Brigard seems to assume this interpretation. He explicitly mentions psychological hedonism as the target of Nozick’s argument on p. 3 of his paper “If You Like It, Does It Matter If It’s Real?” Others including Torbjorn Tansjo seem to read the passage in a similar way. It’s not clear to me that these people would understand psychological hedonism precisely as I have here.
matters to us” suggest that he might be talking about ultimate human motivation. But we can consider the argument in any case.

It its most straightforward formulation, the argument might simply start from the premise that “you would not plug in to the Experience Machine” and it might move swiftly from there to the conclusion that Psychological Hedonism is not true. Thus, the argument might look like this:

First Anti-Psychological Hedonism Argument
1. If you were in the Nozick Scenario you would not plug in so as to get LEM. You would stick with LA.
2. If (1), then you care ultimately about something other than your own pleasures and pains.
3. If you care ultimately about something other than your own pleasures and pains, then PH is false.
4. Therefore, PH is false.

But in this stark form the argument is unpersuasive. Premise (2) is open to question. It could be claimed that your failure to choose LEM is to be explained by the fact that you fear that the machine won’t work; you fear that you would not get the promised pleasures or that you would suffer some unexpected pain while plugged in. This would be bad news if you were unable to escape. Imagine how painful it would be to find yourself stuck for life in a tank of water with electrodes attached to your brain. So it’s possible that you care only about pleasure and pain, and still would not plug in.

We can avoid this problem by stipulating that the subject in question (viz., “you”) is perfectly certain of the outcomes of his choices. Thus, I formulate the argument as follows:

---

Second Anti-Psychological Hedonism Argument

1. If you were in the Nozick Scenario and were certain that LEM would be more pleasant than LA, still you would not plug in so as to get LEM. You would stick with LA.
2. If (1), then you care ultimately about something other than your own pleasures and pains.
3. If you care ultimately about something other than your own pleasures and pains, then PH is false.
4. Therefore, PH is false.

In this form, the argument is a bit more interesting. Your imagined refusal to plug in must be explainable somehow. Even if we try to explain it by appeal to something completely irrational or to a faulty commitment to a strange axiology or an ultimate interest in being moral, that would not reveal a problem with the argument. For it would show that something other than pleasure is functioning as an ultimate motivator. And that would show that PH is false.

In their 1998 book Unto Others: The Evolution and Psychology of Unselfish Behavior Elliott Sober and David Sloan Wilson interpret Nozick’s Experience Machine argument as an attack on psychological hedonism. They claim that the argument fails. It may be interesting to consider their objection.

Early in the book, Sober and Wilson seem to understand psychological hedonism in a standard way: “avoiding pain and attaining pleasure are the only ultimate motives that people have; everything else we want, we want solely as a means to those twin ends.” (S&W, 2) Although Sober and Wilson do not formulate the Experience Machine argument in precisely the way I have formulated it here, their remarks can be taken as a criticism of premise (2).26 They seem to be saying (in effect) that the psychological hedonist can reject premise (2).

---

26 Which says that if you would choose your actual life over the Experience Machine Life, then psychological hedonism is false.
“The hedonist can maintain that \textit{deciding} to plug into the machine is so aversive [painful?] that people almost always make the other choice. When people deliberate about the alternatives, they feel bad [pained?] when they think of the life they’ll lead if they plug into the machine; they feel much better [they have more pleasant feelings?] when they consider the life they’ll lead in the real world if they \textit{decline} to plug in. The idea of life attached to the machine is painful, even though such a life would be quite pleasurable; the idea of real life is pleasurable, even though real life often includes pain. (S&W, 285; bracketed interpretive questions added, FF.)

When they explain why someone might find the idea of life on the machine painful, they proceed to cite the very factors that Nozick mentions – separation from the real world, abandonment of projects and commitments, “unreality”. But S&W cite these as explanations for the pain the subject would feel as he contemplates making the decision to plug in; they claim that a hedonist can cite these things without abandoning his fundamental hedonistic principles. In other words, S&W are saying that it’s consistent with hedonism for the subject to choose the less pleasurable life; it’s consistent with hedonism for him to do this since \textit{the process of choosing} the less pleasurable life is more pleasurable.

It appears to me that S&W have become confused about the theory under consideration. We can clarify this situation by distinguishing between two views. First, there is the familiar form of psychological hedonism that claims that the acquisition of pleasure and the avoidance of pain are our only ultimate motivators; we are always ultimately motivated to perform acts that we believe will maximize our own hedono-doloric balance. This is roughly equivalent to the version of the doctrine that I introduced earlier:

\begin{quote}
PH: Ultimately, each person is motivated solely by a desire to maximize his own hedono-doloric level.
\end{quote}
A different view would maintain that when people are making decisions, they always choose the decision that is more pleasant to make. According to this view, people are motivated not by the pleasures and pains they anticipate receiving as a result of making a certain decision; rather, they are motivated by the pleasantness or painfulness of the decision-making process itself.

PH2: Ultimately, each person is motivated to make a choice only if he thinks the process of making that choice would be maximally pleasant for him.

These two views will give divergent results in any case in which the choice that’s most pleasant to make is different from the choice that will lead to the outcome containing the most pleasure-minus-pain in the long run. Consider a case in which you know you have a tiny dental cavity. You know that a trip to the dentist will be painful in the near term, but on balance would maximize your hedono-doloric balance. You hate pain – especially when it is coming up soon. Hence, the process of making the decision to go to the dentist is painful. We may assume that it is less painful to make the decision to put it off. So PH2 will say that you will be motivated to choose to put off the visit to the dentist.

But since you know that in the long run you will be better off hedonically if you deal with the cavity soon, PH implies that you will be motivated to go to the dentist soon. Hence, if PH is true (and you know the utilities of your options) you will be motivated to go to the dentist soon.

Since PH represents a more familiar sort of psychological hedonism, I assumed that it would make the most likely target for Nozick’s argument. Thus, I formulated the argument so as to have it yield the rejection of PH. S&W also formulate psychological hedonism more-or-less as PH (on p. 2 of their book), and so one might expect them to interpret Nozick’s argument as I have. But then when they get around to discussing that argument (p. 285) it appears that they start thinking of psychological hedonism in a new and unfamiliar way -- as PH2.
The upshot of all this is that if we understand the allegedly Nozickian argument as I have understood it, the remarks of S&W are irrelevant. They seem to respond to the argument by pointing out that it does not refute a different and unfamiliar form of psychological hedonism.

It seems to me that the form of psychological hedonism I have been discussing (epitomized in PH) is probably the most popular form, but it is not the most plausible. It seems to go obviously wrong in cases involving people who have grown weary of sensory pleasure, or who think that sensory pleasure is disgraceful, or who think that there are more important things than “good feelings”. I’d like to sketch what I take to be a more plausible form of psychological hedonism. In order to state this, I need to draw a distinction.

There is an important and familiar distinction between sensory pleasure (and pain), on the one hand, and attitudinal pleasure (and pain) on the other hand. Sensory pleasures are feelings; they have “felt intensities”; we feel them in parts of our bodies; they are in these respects like tickles and itches. They are literally sensations. But a person takes attitudinal pleasure in something when he is pleased about it, or enjoys it, or is delighted concerning it. Attitudinal pleasure is not a sensation. We don’t feel it in parts of our bodies; it is not like an itch. It is a propositional attitude. To see the difference in a stark case, consider a person who has been suffering from a sensory pain for a long time. Suppose he takes a drug that removes the pain. Suppose that the pain is not replaced by any new and pleasant sensation, but just leaves the formerly painful area feeling numb. The subject might take attitudinal pleasure in the fact that he feels no sensory pain. In this case, the subject takes attitudinal pleasure in a certain fact, but does not feel any sensory pleasure. Attitudinal pleasure and pain always have propositional “objects” like that and that’s what makes them propositional attitudes. Sensory pleasures do not have propositional objects.

---

27 For an extended discussion of this distinction, see Chapter 4 of my (2004).
This provides a new way to understand psychological hedonism. We can assume that whenever a person is confronted with a decision, he will consider his options (though in some cases only briefly or at a low level of consciousness). For each option, there will be some amount of attitudinal pleasure that he takes in the prospect of taking that option. We might elicit this by asking, “How much attitudinal pleasure do you take in the idea that you will take this option?” We can take our new form of psychological hedonism to be the view that when people thus face options, and make decisions, they always decide to take the option in which they take the greatest amount of attitudinal pleasure. We can call this PH3.28

It’s important to see that a person could take pleasure in an option even though he knows quite well that the option itself does not contain any pleasure. This comes out most strikingly in decisions concerning things that will happen after death. Suppose I am asked whether I prefer to be buried or to be cremated – after I am dead, of course. I realize that I will feel no sensory pleasure or pain either way, but I might take more attitudinal pleasure in the prospect of being buried than I do in the prospect of being cremated.

With this in hand, let us return to the Experience Machine. Suppose you are given a full and completely persuasive account of what would happen if you were plugged in, and a similar account of what would happen if you were not plugged in. Suppose you were utterly convinced that you would get more sensory pleasure if you were plugged in. Suppose now you are asked whether you want to plug in. Suppose you consider the options. You see that you would get more sensory pleasure if you were to choose to plug in, but at the same time you take less attitudinal pleasure in the prospect of living out the rest of your life in a tank, connected by electrodes. So you say that you would not plug in.

28 It’s conceivable that S&W had something like this in mind when they talked about pleasure taken in the making of a decision.
In this case, Nozick’s reflections on the Experience Machine would not show that the most plausible form of psychological hedonism (viz., PH3) is false. For in the example, as described, if you choose to remain in your actual life, you in fact do choose the option in which you take greater attitudinal pleasure. Oddly, in this case, you take greater attitudinal pleasure in the option that you know will itself contain less sensory pleasure.

7. Summing Up

Some readers note that Nozick’s discussion of the Experience Machine occurs in a context in which he is talking about utilitarianism for animals. These readers think that Nozick used the thought experiment concerning the Experience Machine as part of an argument against utilitarianism. I presented several different possible forms the argument might be thought to take. The first two arguments are completely unpersuasive; the third focuses only on the hedonistic component of hedonistic act utilitarianism. In spite of the facts about context, it seems to me that there is no convincing evidence that Nozick intended to present any such argument. Final verdict: bad arguments; not in the text.

Very many readers seem to think that Nozick was trying to refute ethical hedonism. They think that when Nozick concluded that certain things “matter to us” beyond the things we could get in the Experience Machine, he meant to say that certain things in addition to pleasures have direct impact on our welfare. I presented several arguments, each starting with a premise to the effect that we would not choose life in the Experience Machine and ending with the conclusion that ethical hedonism is false. The arguments differed in the conditions they imposed on the characteristics we are to imagine the subjects having while making the choice. In some cases, it seemed to me, premise (2) was false. In other cases, it was premise (1). None of the arguments seems persuasive. Again, I am not convinced that Nozick intended to present any of these arguments. Final verdict: bad arguments; probably not in the text.²⁹

²⁹ Many readers think that Nozick was not trying to refute just ethical hedonism. Rather, he was trying to refute all mental state theories of welfare. Since he talks broadly about “how experiences feel from the inside” and not narrowly about pleasure, and since he never explicitly mentions hedonism, this interpretation is more plausible than the preceding one. But the arguments would be just as inconclusive. Final verdict: bad arguments; not clearly in the text.
A few readers think that Nozick was trying to refute psychological hedonism. Perhaps when he talks about “what matters to us” he means “what ultimately motivates us”.

Again, the argument could be interpreted in a variety of different ways. Some forms of psychological hedonism might be refuted by the argument so construed – but these forms of psychological hedonism are pretty implausible to start with. A more interesting form of psychological hedonism would remain unrefuted. The textual evidence in support of this interpretation is debatable. Final verdict: unimpressive arguments; not clearly in the text.

My own view concerning the Experience Machine is this: it’s a wonderfully provocative thought experiment; it can provide the basis for lots of interesting and useful discussions in introductory ethics classes. It makes people think about what they really care about. I have no settled view about what Nozick was trying to establish. Perhaps it was nothing more than this: typical human beings care about being in contact with reality, and displaying actual character traits in their actions, and having actual achievements; thus, they care about more than merely having the inner experiences associated with these things. Animals might be like us in these respects. Maybe, therefore, we should not assume that animals care about nothing beyond the quality of their inner experiences. This might have some implications for the notion that some form of utilitarianism accounts for our obligations to animals. I leave it to others to assess the plausibility and significance of the argument so construed.

Appendix

In spite of the fact that Nozick never mentions either intuition or intrinsic value in the Experience Machine passage, some commentators persist in saying that the passage contains an argument that turns essentially on a claim about our intuitions about the intrinsic values of two lives. (Nozick’s own remarks in later writings tend to suggest this sort of view.) Let us consider a typical interpretation of this sort.

---

30 Although it must be acknowledged that it is pretty wild to assume that a cow or a chicken is deeply concerned about the extent to which she is having actual achievements or displaying excellent character traits.
First we must describe the essential features of two possible lives. Following what was said in the text of the paper, let us say that LEM is a marvelously pleasant life in an experience machine. The subject is completely fooled. He thinks he is writing great novels, making lots of friends, and reading interesting books. In fact, however, he is floating in a tank with electrodes connected to his brain. For comparison, we say that LA is a moderately pleasant life in the real world. The subject never writes any great novels, makes only a moderate number of friends, and enjoys reading only a few fairly good books. However, he is in contact with reality, he has actual (modest) achievements, and he does not live in a purely man-made reality.

We stipulate that the lifetime hedono-doloric balance of the person who lives LEM is significantly greater than the lifetime hedono-doloric balance of the person who lives LA. As before, ethical hedonism is understood to be the view that the welfare value of a life is entirely determined by the amounts of sensory pleasure and pain in the life, in such a way that the welfare value of a life is directly proportional to the hedono-doloric balance of that life. The welfare value of a life is understood to be a form of intrinsic value.

We may now state the argument as follows:

1. If ethical hedonism were true, the welfare value of LEM would be greater than the welfare value of LA.
2. But it’s not the case that the welfare value of LEM is greater than the welfare value of LA.
3. Therefore, ethical hedonism is not true.

Let us agree that the defender of this argument attempts to defend premise (2) by saying that it’s all a matter of “moral intuition”. When he reflects with due care on LEM and LA, he can just “see” that the intrinsic welfare value of LEM is lower than the intrinsic welfare value of LA.

I have to admit that I find this latest argument to be fairly interesting. But anyone who is skeptical about moral intuition or intrinsic value would surely find it much less interesting. Among those who have no problem with moral intuition and intrinsic value, the hedonists will claim that they just “see” that (2) is false. So it’s a controversial argument. But there is one point that all readers should agree upon: Nozick did not present this argument in the Experience Machine passage. He never mentioned moral intuition. He never mentioned intrinsic value. These are essential features of the argument presented here, but are not features of Nozick’s argument.

On the other hand, he did say various things about whether his reader would (or should) choose to plug in. Something about choosing to plug in seems to be an essential feature of Nozick’s argument. Yet there is no mention of any such choice anywhere in the argument presented here. Choice simply plays no role in this argument. Thus, the text of AS&U provides no justification for attributing this argument to Nozick. Final verdict on this argument: Possibly an interesting argument; definitely not in the text.

References


De Brigard, Felipe. “If You Like It, Does It Matter If It’s Real?” *Philosophical Psychology*, forthcoming.


