



Open Research Online

Citation

Barber, Alex (2011). Hedonism and the experience machine. *Philosophical Papers*, 40(2) pp. 257–278.

URL

<https://oro.open.ac.uk/29075/>

License

None Specified

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Hedonism and the Experience Machine

Alex Barber

Re-Reading of Robert Nozick, ‘The Experience Machine’, in his *Anarchy, State, and Utopia*, New York: Basic Books, 1974, pages 42-5.*

Introduction

Money isn’t everything, so what is? Many government leaders, social policy theorists and members of the general public have a ready answer: happiness. Philosophers, in contrast, usually see hedonism as a non-starter. This is due in no small part to a sub-section of a chapter in a book by Robert Nozick. ‘The experience machine’ is just four pages long but contains one of the most frequently and favourably cited thought experiments in normative ethics, a gift for anyone lecturing on utilitarianism.

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to the brain. Should you plug into this machine for life, preprogramming your life’s experiences? (1974: 42)

We are expected to follow Nozick in answering ‘No’.

On the surface this is a beautifully simple dismissal of hedonism. Part of its appeal lies in the threat it poses to *all* versions of the claim that pleasure or happiness alone is ‘what matters to us’ (to use Nozick’s phrase). John Stuart Mill may or may not have been more perceptive

* Editors’ Note: ‘Re-Readings’ is a regular feature in *Philosophical Papers*. Authors are invited to write on a past article, book, or book chapter that they deem, for whatever reason, to deserve renewed attention. Authors are encouraged, where appropriate, to discuss the work’s reception by and influence upon the philosophical community.

than Jeremy Bentham about the relative merits of what he called the higher and lower pleasures, but Nozick's example challenges any theory of value that refuses to look beyond subjective mental states.¹

The example has become far more familiar than Nozick's original text. Given the frequency with which philosophers continue to endorse it, and the relative lack of critical scrutiny, it is worth asking whether the underlying argument is as successful as it is elegant and influential.² There are other reasons for taking a fresh look. The philosophical and non-philosophical worlds have changed in the thirty-seven years since the example was first published. Thought-experiment methodology, in which often wildly improbable scenarios are used to eke out our pre-theoretical intuitions, has slowly fallen out of favour. At the very least, philosophers are more wary of its use than they once were. In another development, happiness has enjoyed a resurgence among social and economic theorists. Subjective happiness is increasingly being favoured over financial wealth as a measure of a society's well-being, and by implication as an indicator of good policy. Psychologists have also made progress in understanding the factors that affect levels of subjective well-being. And finally, virtual games that in some ways resemble Nozick's imaginary machine have grown increasingly

1 I am following a convention in the literature of using 'hedonism' to be about happiness and not just pleasure in some narrower sense, as etymology would require. I am, though, treating happiness as a subjective mental state—and specifically one that supervenes on narrow physical properties of the kind that Nozick's neuroscientist can manipulate so skilfully.

2 A small sample of favourable mentions: Griffin 1986: 9, Brink 1989: 223-4, Becker 1992: 25, Darwall 1997: 162, Hooker 2000: 39, S. Bok 2010: 24-8, and Hurka 2011: 68-70. Silverstein 2000 gives a more extensive list, including several textbooks (p. 282n). Among the critics, who are far fewer in number, Donner 1991: 69-78 rejects the example but only because she embraces a non-subjectivist account of happiness. Brandt 1996: 42 is sceptical of the example but misses the important point that you would be unaware you were in a mere machine. Sumner 1996: 94-6 raises some problems for the argument but does not consider in detail whether it might be modified, and ends up endorsing it anyway (p. 98). Tännsjö 2007 thinks it works except against narrow Benthamite hedonism, which he supports. Sober & Wilson 1998: 285 reject the example's use against motivational/psychological hedonism. Silverstein 2000, Rivera-López 2007, and Brigard 2010 offer criticisms that are consonant with some of my own (see notes 7, 14, 15).

sophisticated and more widely consumed. One gaming company uses the slogan ‘Step inside’.³

My contention in this paper is that the argument Nozick generates from his example is riddled with difficulties. This is not because I think he is wrong to reject hedonism, or even that his reasons for doing so are mistaken. It is just that the experience-machine example fails to capture these reasons as effectively as he and many others assume. After spelling out his argument and then my reservations about it, I show how a case against hedonism can be made more effectively with a different argumentative strategy and different examples.

Nozick’s Argument

So what is Nozick’s argument? There are, it turns out, two different ways of reading the text of the 1974 presentation, though a later discussion of the same example helps us to see which he had in mind and why.

The 1974 presentation opens with Nozick describing the experience machine, as quoted already. He assumes that a hedonist would plug in but that you would not plug in. He concludes:

We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it. (1974: 44)

In other words, your response shows that hedonism is mistaken—though he never uses that label, instead characterising the view he opposes as a theory of value according to which ‘[nothing] matters other than how people’s experiences feel “from the inside”’ (p. 42).

Other things Nozick says add flesh to this skeleton of the argument. He makes a number of fine-tunings—five in total—all apparently

3 For scepticism over the use of intuitions about thought-experiments in ethics, see (for example) Horowitz 1998. For more and less zealous application of happiness studies to social policy see (for example) Layard 2006, D. Bok 2010, and Stiglitz, Sen & Fitoussi 2011, the latter a report commissioned by French President Sarkozy. The company inviting you to ‘Step inside’ is Xbox.

designed to ensure that if you were a genuine hedonist you would have no grounds for not entering. First, in case you were worried about missing out on life's possibilities, he envisages a smorgasbord of pre-programmable experiences based on research into real lives. Second, presumably to see off alarm you may experience over the finality of your decision, he allows a biannual unplugging so you can review your choice. Third, even hedonists would resist plugging into the machine if they knew that, having done so, they would know it was a grand illusion. That could feel like prison, as several Hollywood films have since brought out.⁴ Knowing that one's experiences are inauthentic undermines the happiness one might otherwise feel. This appears to be why Nozick writes that 'while in the tank ... you'll think it's all actually happening' (p. 43). Fourth, you may have moral reservations owing to obligations you feel towards others. Nozick helps you to set these aside by making the experience-machine option available to everyone. And finally, in case you suffer anxiety at the prospect of waving goodbye to your current life, he stresses how puny this anxiety is compared to the happiness that will ensue.

Thus far, Nozick's argument is reasonably clear. The complication arises in connection with three anti-hedonistic judgements he makes about what matters to us, each of which accords with our unease at the prospect of an experience-machine life (pp. 43-4). The first is that, when we want the experience of doing something, we often want that experience as the by-product of the doing, not for its own sake. Wanting to do is therefore often prior to wanting the experience of doing. His second judgement is that we want to be a certain way. In particular we do not want to be indeterminate blobs in tanks. A blob in a tank is not courageous, kind, witty, loving, or intelligent. 'Plugging into the machine is a kind of suicide', says Nozick, an erasing not only of our earlier selves but also of selfhood of any kind. And finally, we value the possibility of coming into contact with and exploring a reality that is deeper than the subjective one.

4 Most notably, *The Truman Show* (1998) and *The Matrix* (1999).

These three anti-hedonistic judgements generate an ambiguity in the logic of his argument. Nozick could be saying: 'Hedonists say we should get in; here are three reasons we shouldn't; ergo here are three reasons why hedonism is wrong'. The reader's intuition, 'I wouldn't get in', would play no substantive role in the argument on this reading. The example itself, on this construal, is just one possible and vivid way to bring out the contrary implications of hedonism and the three judgements about what matters to us. The three judgements themselves bear the argumentative weight.⁵

On my preferred reading of his argument the decision we make is not an optional extra that helps us appreciate his reasoning. In brief, this version of the argument runs: 'Hedonists would plug in, you would not, ergo hedonism is wrong as a theory of what matters to you'. The three anti-hedonistic judgements function as conjectures about where hedonism has gone wrong, but are not part of his argument *that* it has gone wrong. Should they turn out to be misdiagnoses, hedonists would still have a case to answer.

There are clues in the original article as to which is the intended reading (e.g., the passage from p. 44 quoted at the start of this section), but in a later commentary on his example, Nozick is quite explicit:

[T]he machine example must be looked at on its own; to answer the question by filtering it through a fixed view that internal experience are the only things that *can* matter (so of course it would be all right to plug into the machine) would lose the opportunity to test that view independently. One way to determine if a view is inadequate is to check its consequences in particular cases, sometimes extreme ones, but if someone always decided what the result should be in any case by *applying* the given view itself, this would preclude discovering it did not correctly fit the case. (Nozick 1989: 105)

He is warning against filtering one's response through hedonism; but by implication he is committed to not filtering it through anti-hedonism

⁵ This reading seems to be followed by, for example, Griffin 1986: 9. Most commentators fail to explain how they see Nozick's three judgements as fitting into his argument.

either. It is the immediate inclination that matters. This is what I referred to earlier as ‘thought-experiment methodology’. Indeed it is quite a hard-nosed version of the methodology. Most appeals to thought experiments are made with the aim of achieving consistency between one’s theory and one’s *reflective* judgements, not one’s theory and one’s first impulse. But the main point I want to make here is interpretive: the response we give to the prospect of getting into the experience machine is integral to the argument against hedonism, whereas the three anti-hedonistic judgements are attempts to explain where hedonism has gone wrong. If the example needs to be buttressed with anti-hedonistic claims, it has failed.

Problems with Nozick’s Argument

It will help if the different elements of the preceding section are mustered into a structured argument. The objections I am going to raise all relate to one or other step in the following:

- | | |
|-----------------|---|
| Premise 1 | You and other readers would not plug in to the experience machine. |
| Premise 2 | A genuine hedonist would plug in to the experience machine. |
| Subconclusion | Neither you nor other readers are genuine hedonists. |
| Premise 3 | If neither you nor other readers are genuine hedonists, hedonism is mistaken. |
| Main Conclusion | Hedonism is mistaken. |

This is meant to be a plausible starting representation of Nozick’s argument. Some of the objections I raise will force modifications that could have equal claim to be what he had in mind all along. In any case, the argument fails with or without the modifications.

The first premise is the basic datum Nozick extracts from his example.⁶ This ought to be straightforward but is not. Many students respond to the example the wrong way when first presented with it in moral philosophy lectures. I will call this the *enterers problem*.⁷ As virtual technology makes the flight from reality ever more familiar and unthreatening, we can expect the percentage of yea-sayers to increase. Moreover, there are dialectical reasons for making the prospect of getting in less unappealing (see the discussion of the first inference, below).

These deviants are a problem because we want there to be an objective answer as to whether hedonism is correct as a theory of value. Nozick seems to think that very few will opt to plug in. He addresses would-be enterers only to say that they should ask whether their first impulse was not to plug in (1989: 105). But this is hardly satisfactory. It does not help with the data, which do seem to capture first impulses, but in any case first impulses are often a poorer guide to our ‘real’ opinions than considered responses.

Elsewhere, Nozick provides the seeds of a more promising way of dealing with would-be enterers. He asks ‘Why feel any distress at all if your decision *is* the right one?’ (1974: 43). From this we can see the problem for hedonism in a more nuanced light: it cannot explain even the smallest tug away from the desire we all have for maximal happiness. Even if we do not all think we should resist the machine’s temptations, we nearly all feel *some* resistance to jumping in, and that resistance alone is enough to show that we are not hedonists. Genuine hedonists say the

6 Nozick wavers between ‘would’ (1974: 44; 1989: 105) and ‘judge that you should’ (1974: 42). As I work through the argument I will consider revisions to the premises that reflect this wavering, but I read him as being interested in your views about what you (self-interestedly) *should* do, but what you *would* do is our best guide to these views.

7 On a small sample audience of philosophy undergraduates I found the split to be, quite surprisingly, 52% non-enterers to 48% enterers (but see the next note). Brigard 2010 provides more credible data and stresses that results vary according to how the example is described to participants. One of his most telling findings is that participants are rather less anxious to leave a simulated reality than they are to stay out of one—what could be called the non-exiters problem.

answer is obvious and would get in; we think either that the answer is non-obvious or that the answer is obviously the opposite. So we cannot be hedonists.⁸

This approach to the enterers problem has important ramifications that I will return to later. For now, I want to move on to the second premise. This says that a genuine (as opposed to self-proclaimed) hedonist, valuing only 'how things feels from the inside', would plug in, a seemingly uncontroversial claim. And Nozick is surely right that a crude Benthamite sensation hedonist—someone for whom the value of a mental state comes down to a 'raw feel' that is independent of its semantic content—would get in. But if that were his only target, Nozick's argument would have limited interest. He needs to show that more sophisticated forms of hedonism also conform to this second premise.⁹

The kind of hedonism I have in mind, which I will call preference hedonism, lies between sensation hedonism and a preference-satisfaction theory. A preference-satisfaction theorist would not plug into the machine, since doing so would not lead to the satisfaction of their preferences, only to the appearance of satisfaction. This is not a problem for Nozick since preference-satisfaction theories are not hedonistic and so are not a part of his target. Preferences may be mental states, but whether they are satisfied depends on features of the non-mental world.

According to preference hedonism what matters is not your preferences being satisfied, but your thinking they are or will be. This sets it apart from a preference-satisfaction view as a species of hedonism

8 When I changed the question put to the audience mentioned in the previous note, the effect was striking despite the small sample size and non-rigorous conditions. The split went from 52% of non-enterers versus 48% of non-enterers to 87% saying it was not obvious they should enter versus 14% saying this was obvious. (The 87% was made up of 23% who thought it obvious they should stay unplugged plus 64% who thought neither option was obvious.). The absolute figures may be unreliable but the change looks significant.

9 Tännsjö 2007 is unusual in being content to defend Benthamite hedonism against Nozick and to leave broader conceptions to their fate.

in the broad sense: value is being treated as all about subjective mental states. But unlike sensation hedonism it sees significance in the fact that the valuable mental states are *about* things. Many of our happiness-relevant preferences have rich semantic content, in that (unlike raw sensations) there is a distinction to be drawn between thinking they are satisfied and their being satisfied; thinking these preferences are satisfied is an important source of happiness. In Mill's words:

[T]here is no known Epicurean theory of life which does not assign to the pleasures of the intellect, of the feelings and imagination, and of the moral sentiments, a much higher value as pleasures than to those of mere sensation. (1863: 11)

We could, of course, broaden Mill's list of content-rich, happiness-inducing mental states.

It is easy to suppose that preference hedonists, unlike preference satisfaction theorists, would plug in to the machine. After all, this would give them exactly what they say is in their interests. And since this is a form of hedonism, Nozick would get what he was after. But on reflection a preference hedonist's reaction is hard to predict. Suppose I want to write a groundbreaking novel or have profound friendships. I know the machine is capable of creating in me the impression that these wants are met. But the content of my preference is not that I want *the impression* of good company or of writing a great novel. Even if I am a preference hedonist, I am also someone with preferences of my own, preferences with a content rich enough to make demands on the real world, preferences that therefore militate against my getting in. True, the theoretical belief about what is good for me might itself give rise to a preference that wins out over my other preferences, leading me to plug in; but it might not. At the very least, preference hedonists would be torn, and we have already seen that Nozick needs their decision to be obvious in order to overcome the enterers problem.¹⁰

10 Rivera-López 2007: 79-84 looks in some detail at the question of whether a mental state welfarist can consistently refuse to enter the machine given their real-world preferences. He argues that they can. My view is that they are conflicted. (See also note 18 below.)

Nozick should perhaps have noticed this possibility himself when he reflected on what to say to you if your reason for not plugging in is that you have obligations to, say, ageing parents, and so not necessarily that you are not a genuine hedonist. His response—the fourth of the five ‘fine-tunings’ recounted in Section 2 above—was to reassure you that they will be cared for, albeit by having experience machines of their own. In other words, your moral preference, if we can call it that, would be met *in reality*. He can hardly take this line for all your preferences about the real world. And yet a preference to raise children, or to collect stamps, or to become leader of the UN is just as capable of deterring you from plugging in.

I think this forces Nozick to change Premise 2:

P2 (revised): Hedonism entails that you should plug in.

This revised premise is plausible even of preference hedonism. On the other hand, it pushes the spotlight onto the acceptability of the inference that follows, from P1 and this new P2 to the subconclusion. Prior to the change, the inference was a straightforward deductively valid inference. It is now an inference to the best explanation. Specifically, the best explanation of your response in P1, given its inconsistency with what hedonism recommends, is that you do not accept hedonism as an account of value. Or rather, you do not, in your heart of hearts, or at a gut level, accept it, even if you profess it at a theoretical level. Nozick, here, is like a skilled cross-examiner, getting us (the witness) to reveal what we really think by answering spontaneously and so missing the ramifications of our reply for our philosophical opinions (analogous to a concocted story). He achieves this by asking us what we would do in this strange setting. Our hypothetical actions speak louder than our theoretical words.

So how good is this new inference? As with any inference to the purportedly best explanation, this one must see off competitor explanations. And there do seem to be other ways of accounting for our response to the example that are compatible with our being (even in our

heart of hearts) hedonists. Perhaps plugging into the machine strikes us as simply too terrifying or claustrophobic a prospect even if we know that it is the right thing for us. Some refuse to fly but will drive even if they know this is far more dangerous, and boarding an airplane is as nothing compared to volunteering oneself up to a different—and delusional—plane of reality.¹¹

Fear is not the only potential confounding factor. Thinking of oneself isolated in a tank, cut off from genuine interaction with family or friends, evokes a sense of loneliness. Moreover, humans are naturally risk averse, especially in situations that, as here, involve large and unfamiliar changes. Fear, loneliness and risk aversion could all be regarded as manifestations of our natural rationality deficit, perhaps—of the kind that cognitive psychologists have become increasingly interested in over recent decades. That is certainly how a hedonist would see them. But this hardly disqualifies them from being alternative sources of our reluctance to embrace the virtual option.

There are changes Nozick could make to help his case, particularly by ruling out irrational resistance. The idea of plugging yourself into a machine has echoes of being in hospital, on a dialysis machine for example. So he could note that we all already own and enjoy using something like an experience machine on a daily basis: our beds. Couldn't there be simple, non-intrusive ways of sleeping for indefinitely long periods after having been trained to have only fabulous dreams? Another worry we may have is that we would choose the wrong thing to experience (like: finally understanding the Preface in Hegel's *Phenomenology of Spirit*). This corresponds to a fairly robust datum from—indeed it is the *raison d'être* of—Happiness Studies: we are not that good as individuals at telling what will make us happy. Nozick's smorgasbord does not quite get around this problem because it still leaves the choice to us. So why not just let the machine make the choice for us? Another

¹¹ Others make this and related points, e.g., Sumners 1996: 94-5. For a contrary view, see Silverstein 2000: 284-5.

less robust datum from Happiness Studies is that we are all stuck on a kind of ‘hedonic treadmill’.¹² When our circumstances change we experience temporary leaps or dips of happiness and then drift back to a predefined level, albeit one that varies from individual to individual. Neurological intervention could prevent that happening to us in the machine, and periodic injections of melancholy or set backs would see off any concern that a life of relentless joy might start to wear thin. Nozick himself makes a suggestion in the 1989 presentation that could help to deal with the worry about loneliness: we could all be participating in a co-ordinated virtual experience, though of course we would not know anything about this set up once we were inside the machine.¹³

How successfully these or other tweaks can deal with all confounding factors is unclear.¹⁴ One problem emerging from the suggestions just given is that as each possible factor is ruled out, plugging in becomes less repellent. Many present-day computer-generated virtual realities are enjoyed precisely because they allow multiple players, so introducing a communal element to the experience machine could make the prospect quite appealing (as Nozick admits). Letting the machine’s computer work out what will make for a good life could introduce an attractive element of surprise. And so on for the other changes, with the result that the enterer’s problem becomes more pronounced. I will soon question what I earlier took to be Nozick’s most promising response to that problem.

So far I have been highlighting alternative possible explanations of our reluctance to plug in. But Nozick’s own explanation can also be faulted, independently of the attractiveness of competing ones. We saw that a hedonist would not necessarily plug in. This is what forced the

12 The phrase was coined in the early 1970s and the thesis most famously supported by a study of lottery winners and accident victims (Brickman, Coates, & Janoff-Bulman 1978). Significant reservations are expressed in Deiner, Lucas, & Scollon 2006.

13 1989, p. 107.

14 The possibility that we are simply manifesting the status quo bias, something we do in plenty of other contexts, is examined—and supported empirically—in Brigard 2010.

revision to P2 in the first place. But if a hedonist would not get in, your not being a hedonist would hardly explain your not getting in. If sea levels would have risen just after the middle Miocene era even if the polar ice caps had not melted, sea levels having risen then is not best explained by polar ice caps having melted then, other things equal. More generally, other things being equal, E is unattractive as an explanation of P if the probability of P is high even on the assumption that not-E. No fine-tuning of the example can rid Nozick of this particular difficulty since it depends only on the hedonist knowing that her happiness-relevant preferences would not be met in the machine.

Would it help to change the first premise in the same way that P2 was changed?

P1 (revised): Your heart-of-hearts theory of value entails that you should not plug in.

This and the earlier revision to P2 effectively restore the inference's deductively valid status, bypassing the need to infer by appeal to the best explanation. But this change really just buries the problem. Unless we already make non-hedonistic—and so question-begging—assumptions about the content of our heart-of-hearts theory (by taking them to include one or more of Nozick's three anti-hedonistic judgements, for instance), the only grounds for this revised premise is that your decision not to get in is a manifestation of your commitment to some heart-of-hearts theory rather than to, e.g., fear, loneliness, etc. So, alternative explanations of your decision are still a threat.

Let us turn now to the second half of the argument, the inference from the subconclusion via P3 to the main conclusion. Nozick does need to make this journey. It is not enough for his purposes that, when push comes to shove, we *think* hedonism is mistaken, since we could be mistaken that it is mistaken. Nozick says nothing explicit in defence of this inference in either the original 1974 discussion or the 1989 commentary. That could lead us to conclude that he was simply relying on the assumption, common at the time though somewhat less so now,

that our core beliefs, the intuitions revealed by examples like this one, are the best guide there is to philosophical, and in particular ethical, truth.

I suspect there is more going on than simple intuition mongering. One phrase crops up over and over in both pieces: ‘what matters to us’. Nozick can be read as making the limited claim that when it comes specifically to what matters to us, our intuitions, our core beliefs, *are* the best guide we have. This need not commit him wholeheartedly to the more general and increasingly disreputable trade in intuitions. P3, if fleshed out, would be saying:

P3 (revised): If you are not a genuine hedonist (about what matters to you), hedonism is mistaken (as a theory of what matters to you).

But this key phrase is highly ambiguous. The ‘to you’, in particular, can mean ‘as you see it’; and it can mean ‘modulo you’, or, ‘setting the concerns of others to one side’. In the first sense, what matters to you most certainly is equivalent to what you think matters, since this is true by definition. But you could very well be mistaken about what matters to you in the second sense, i.e., about what is in your interests. Hedonists have, historically, favoured social and ideological reform. For them, a discrepancy between what people judge to be best for them and what is in fact best for them is hardly the deathblow Nozick appears to assume; his use of the refreshingly non-technical but in reality quite ambiguous phrase, ‘what matters to us’, should not be allowed to disguise this fact.¹⁵

This ambiguity is not the only problem with the final step in his argument. To see this, recall the enterer’s problem first raised in connection with P1. The solution I offered on Nozick’s behalf was to change the datum from ‘You would not plug in’, to ‘You do not find it obvious that you should plug in’. But the most we can get from this plus

¹⁵ Though he discerns it in the word ‘matters’ rather than ‘to us’, Silverstein 2000: 285-6 insists on something like this ambiguity, and uses it against Nozick in part of his defence of hedonism.

any suitable wording of P2 is the far weaker subconclusion, that *you do not find hedonism obvious*. This, after all, is a perfectly good explanation of your not finding it obvious that you should plug in. There is no need for you to think hedonism is false, let alone obviously false, to be troubled over what to do. Nozick's error is manifest in a sentence quoted earlier: 'Why feel any distress at all if your decision *is* the right one?' (1974: 43). You could think your decision is the right one but feel distress because you are not *sure* it is. But if the sub-conclusion is weakened in this way, the inference in the second half of the argument is even weaker than I suggested earlier. P3 would need to become:

P3 (revised): If neither you nor other readers find hedonism obvious, hedonism is mistaken.

This itself is hardly obvious.¹⁶

Salvaging Nozick's Criticism

In the previous section I presented four objections to Nozick's argument. Ignoring their occasionally subtle interplay, these are:

1. Many find the option of plugging in more tempting than Nozick allows.
2. There are different kinds of hedonist, and some would not plug in.
3. That we are non-hedonists is not the best way to explain our not plugging in.
4. The inference from what we intuit about what is in our interests to what is genuinely in our interests is shaky.

I want to end by sketching a strategy for arguing against hedonism that is immune to these difficulties yet still takes its cue from Nozick's

¹⁶ I have here adapted the strategy used by Lenman 2004 to defend hedonism's stablemate, consequentialism, against a famous thought experiment in Williams (1973: 108-18) involving Jim and twenty villagers who will be shot by Pedro unless Jim shoots just one of them.

discussion. One difference is that I will use examples that do not involve the bizarre, life-changing and alienating decision Nozick presents us with. More significantly, I won't argue that hedonism is wrong simply because it is counterintuitive. Instead, my goal will be to show that hedonism—specifically preference hedonism—cannot be maintained consistently.

The strategy is to use a tension within preference hedonism against it rather than in its favour. We saw that the preferences of a preference hedonist can clash with their hedonism. In Nozick's argument, this internal discord helped to save the view. The original P2 had to be rejected since preference hedonists would at the very least be torn over whether or not to plug in; and for the same reason the subconclusion was a poor explanation of P1, undermining the first inference. What we need is a way of converting this tension within preference hedonism from an escape route for it into an objection to it.

The examples I want to use are built out of the notion, drawn on already, of a preference that is *happiness-relevant*: your being under the impression that it is satisfied leaves you happier. Nozick's examples were of making friends and writing great novels but they could be any preferences at all. For any happiness-relevant preference, it is usually easy enough to imagine circumstances in which you think it is satisfied but then discover it was not—and to do so without positing an experience machine.¹⁷ You find out that the 'homemade' cookies you just ate were really baked from store-bought dough; that they have been letting you win at poker; that your lover in fact hates you; that the wilderness you have made your home contains the world's largest call centre, cunningly disguised by artificial trees; that your best friend is paid out of a trust-fund set up by your parents; that the children you have taken hard-won but profound joy from helping to raise are in fact remotely-controlled robots; and so on. In each case you are happy until

17 Several authors note that non-fantastical settings can be used to make Nozick's point, e.g., Griffin 1986: 9 and Sumner 1996: 96-7.

you discover what a chump you have been, at which point your happiness evaporates.

A chump (in this context) is someone who is completely happy, unaware it's all massively phoney. Is it better to be a chump, happy but deceived, or a non-chump, unhappy but in the know? We could use Nozick's example to pose this same question, but doing so brings with it plenty of unhelpful complications. All we need for an example of chumpery are two simple ingredients: a happiness-relevant preference and a situation in which you believe the preference is satisfied even though it is not. Moreover, Nozick asks the question in order to elicit a negative answer, and thereby to show that we find hedonism counterintuitive. I am not invested in showing this, and instead ask the question so as to bring out the tension within hedonism: a hedonist has no coherent response to the question. Their happiness-relevant preferences mean they want to be a non-chump; their hedonism means they want to be a chump. Asked whether it is better to be a chump or not a chump, a hedonist has no answer to give.

The tension gives rise to more than a simple clash of wants, the preference itself and a desire to be happy. If that was all there was to it, the objection would be hopelessly weak. Rather, it is a clash between values. One is the hedonistic value of happiness; the other is the value that is integral to the preference for friendship over pseudo-friendship, or whatever it may be. The first of these, moreover, is meant to be an exclusive and comprehensive value, and so cannot logically co-habit with the value of genuine friendship.

The trouble with preference hedonism, then, is that it presupposes the existence of happiness-relevant preferences but at the same time rejects the values that are integral to these preferences. It is because we value the contribution we are making towards our children's welfare that we take pleasure from fulfilling a parental role; it is because we value (to the extent of being willing to pay extra for) cheese that comes from cows that graze on Alpine flowers rather than a chemical additive that we are happy when we think this is what we are eating. These are not everyone's

preferences of course. Many take no pleasure from children or thoughtfully sourced cheese. But the argument works for *any* happiness-relevant preference, so long as it is possible to be deceived about that preference being satisfied.

I said I would sketch an experience-machine free argument against hedonism, and that is all I claim to have done, but I will end by anticipating two possible responses. This first would be for a defender of preference hedonism to say: ‘When I am being a human being I value my preferences being satisfied in reality; but when I am being a philosopher I am a hedonist and care only that I think they are satisfied. Right now I am being a philosopher.’ This sounds a little like something Hume would say whenever he saw the finger of reason pointing where his human nature would not allow him to travel. But Hume’s position is in fact more nuanced. He does not assume there are two selves or perspectives within us that can be teased apart and maintained consistently. He writes ‘Be a philosopher; but, amidst all your philosophy, be still a man’ (*Enquiry Concerning Human Understanding*, Section I). In other words, do not let your philosophy float free from your humanity. It is in this spirit that I put the objection in the form of a question: how would *you* respond to the question of whether it is better to be a chump or a non-chump. Appealing to the existence of two perspectives on value no more gets around the inconsistency I have diagnosed than talking of an internal perspective from which we have freewill and an external perspective from which our actions are governed by physical laws suffices in itself to establish compatibilism in the freewill-determinism debate.¹⁸

A different response is to see the value embedded in our preferences as instrumental and the value of the happiness arising from the

18 In an intriguing and rich discussion of this issue, Rivera-López 2007 talks of distinct but compatible levels, points of view, or perspectives on what is valuable (p. 84), and confronts the threat of what he calls ‘schizophrenia’ directly (pp. 85-7). His position deserves closer attention but my sense is that he manages to show only that, in his terminology, mental-state welfarists can look favourably on the adoption of reality-centred desires; he fails to show how one can consistently commit to the values implicit in both.

successful pursuit of these preferences as intrinsic. This promises to let both values cohabit within a single individual. But like the previous response it requires a split in the mind of anyone adopting it. The content of our happiness-relevant desires has something of the categorical about it. Suppose you photograph, after decades of trying, an extraordinary undersea occurrence. The value implicit in the preference that led to your pursuit of the photograph does not strike you as instrumental, and it would take some athletic double-thinking to, at one and the same time, value photographing a real event (as opposed to one staged by friends) and regard the whole episode as valuable merely for the happiness it brings about.¹⁹

Summary

Nozick's experience-machine argument relies too heavily on drawing out our anti-hedonistic intuitions about what is prudentially valuable, and in any case fails to demonstrate that our intuitions really are anti-hedonistic. I have sketched what I think is a more promising approach, albeit one in the spirit of Nozick's position. This involved looking at scenarios where the deception is not all encompassing, but the more important difference was in strategy. Instead of framing the case against hedonism as a battle between our intuitions and a philosophical theory, I tried to show that hedonists—and specifically preference hedonists—are committed to an unsustainable clash between the value they place on

19 Kawall 1999 appears to invoke an intrinsic/instrumental distinction but his position is complicated because he also seems interested in defending a hedonistic theory of well-being that falls short of being a hedonistic theory of value, as when he acknowledges 'rival values to our own well-being', such as, say, the value of having a genuine rare stamp rather than a good forgery (pp. 385-6). And Silverstein 2000: 293 quotes Railton 1989: 169 favourably as saying that 'if our diverse intuitions about what is prudentially valuable "can be explained in a unified way by invoking a substantive conception of happiness ... then the hedonist can claim that, *despite appearances*, these other ends owe their hold upon us to the role they have played in the creation of happiness.'" The phrase I have italicised confirms the need for an element of self-deception on this approach; and explaining the hold of our ordinary preferences on us is distinct from showing that the values within them can be maintained consistently alongside a hedonistic conception of value.

happiness and the values built into the preferences even they recognize as key to human happiness.

My quarrel has been with Nozick's example, not with what he tried to show with it. He too, elsewhere, appeals to simple, everyday cases (1989: 106). He also sees the importance of distinguishing, and the dangers of conflating, the value implicit in a desire to do something and the value of the happiness that arises from satisfying that desire. Echoing Aristotle he writes:

We want to *do* certain things, and not just have the experience of doing them. In the case of certain experiences, it is only because ... we want to do the actions that we want the experiences of doing or thinking we've done them. (1974: 43)

As social policy becomes increasingly targeted at promoting happiness instead of the more traditional monetary values that may enable us to engage for real in worthwhile activities, it is an important message.²⁰

Open University, United Kingdom
a.barber@open.ac.uk

References

- Becker, L.C. 1992. 'Good Lives: Prolegomena', in E.F. Paul, F.D. Miller, Jr., & J. Paul, eds., *The Good Life and the Human Good*, Cambridge: Cambridge University Press. 15-37.
- Bok, D. 2010. *The Politics of Happiness: What Government Can Learn from the New Research on Well-Being*. Princeton: Princeton University Press.
- Bok, S. 2010. *Exploring Happiness: From Aristotle to Brain Science*. New Haven: Yale University Press.
- Brandt, R.B. 1996. *Facts, Values and Morality*. Cambridge: Cambridge University Press.

²⁰ I am grateful to audiences at the University of Northampton and at an Open University Residential School in Bath for their questions and comments, as well as to Chris Belshaw, Timothy Chappell, Lindsey Porter, Robin Scaife and the journal editor for questions and helpful suggestions.

- Brickman, P., Coates, D., and Janoff-Bulman, R. 1978. 'Lottery Winners and Accident Victims: is Happiness Relative?' *Journal of Personality and Social Psychology*, 36(8): 917-27.
- Brigard, F. de, 2010. 'If you like it, does it matter if it's real?' *Philosophical Psychology*, 23(1): 43-57.
- Brink, D. 1989. *Moral Realism and Foundations of Ethics*. Cambridge: Cambridge University Press.
- Darwall, S. 1997. 'Self-Interest and Self-Concern'. *Social Philosophy and Policy*, 14(1): 158-78.
- Deiner, E., Lucas, R.E., & Scollon, C.N. 2006. 'Beyond the Hedonic Treadmill: Revising the Adaption Theory of Well-being'. *American Psychologist*, 61(4): 305-14.
- Donner, W. 1991. *The Liberal Self: John Stuart Mill's Moral and Political Philosophy*. Ithaca: Cornell University Press.
- Griffin, J. 1986. *Well-being: Its Meaning, Measurement and Moral Importance*. Oxford: Clarendon.
- Hooker, B. 2000. *Ideal Code, Real World: a Rule-Consequentialist Theory of Morality*. Oxford: Clarendon.
- Horowitz, T. 1998. 'Philosophical Intuitions and Psychological Theory'. *Ethics*, 108(2): 367-85.
- Hurka, T. 2011. *The Best Things in Life: A Guide to What Really Matters*. New York: Oxford University Press.
- Kawall, J. 1999. 'The Experience Machine and Mental State Theories of Well-Being'. *Journal of Value Inquiry*, 33: 381-87.
- Layard, R. 2006. *Happiness: Lessons from a New Science*. 2nd edition. London: Penguin.
- Lenman, J. 2004. 'Utilitarianism and Obviousness'. *Utilitas*, 16: 322-325.
- Mill, J.S. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.
- Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Nozick, R. 1989. 'Happiness' in his *The Examined Life: Philosophical Meditations*, New York: Simon and Shuster. 99-117.
- Railton, P. 1989. 'Naturalism and Prescriptivity'. *Social Philosophy and Policy*, 7: 151-74.

- Rivera-López, E. 2007. 'Are Mental State Welfarism and Our Concern for Non-Experiential Goals Incompatible?' *Pacific Philosophical Quarterly*, 88: 74-91.
- Silverstein, M. 2000. 'In Defense of Happiness: A Response to the Experience Machine'. *Social Theory and Practice*, 26(2): 279-300.
- Sober, E. and Wilson, D. S. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass.: Harvard University Press.
- Stiglitz, J.E., Sen, A. & Fitoussi, J-P. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Later published (with a foreword by President Sarkozy) as *Mismeasuring our Lives: Why GDP Don't Add Up*. New York: New Books, 2010.
- Sumner, L.W. 1996. *Welfare, Happiness and Ethics*. Oxford: Clarendon.
- Tännsjö, T. 2007. 'Narrow Hedonism'. *Journal of Happiness Studies*, 8(1): 79-98.
- Williams, B. 1973. 'A Critique of Utilitarianism'. In J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.