How to Use the Experience Machine

EDEN LIN

Rutgers University, Newark

The experience machine was traditionally thought to refute hedonism about welfare. In recent years, however, the tide has turned: many philosophers have argued not merely that the experience machine doesn't rule out hedonism, but that it doesn't count against it at all. I argue for a moderate position between those two extremes: although the experience machine doesn't decisively rule out hedonism, it provides us with some reason to reject it. I also argue for a particular way of using the experience machine to argue against hedonism – one that appeals directly to intuitions about the welfare values of experientially identical lives rather than to claims about what we value or claims about whether we would, or should, plug into the machine. The two issues are connected: the conviction that the experience machine leaves hedonism unscathed is partly due to neglect of the best way to use the experience machine.

Nozick's experience machine was traditionally thought to yield a decisive refutation of hedonism about welfare. In recent years, however, the tide has turned: many philosophers have argued not merely that the experience machine doesn't decisively rule out hedonism, but that it doesn't count against it at all. In their view, the experience machine doesn't even give us *pro tanto* reasons to reject hedonism: it provides no evidence that hedonism is false. Hedonism used to be dismissed with a brief mention of the experience machine. But nowadays, many are equally dismissive of any attempt to invoke the experience machine against hedonism.

In this article, I argue for a moderate position between those two extremes: although the experience machine doesn't decisively rule out hedonism, it gives us *pro tanto* reasons to reject it. I also argue for a particular way of using the experience machine to argue against hedonism – one that appeals directly to intuitions about the welfare values of experientially identical lives, rather than to claims about what we value or claims about whether we would, or should, plug into the machine. As I will explain, the two issues are connected: the conviction that the experience machine leaves hedonism unscathed is partly due to neglect of the best way to use the experience machine.

RECENT DISMISSALS OF THE EXPERIENCE MACHINE

Recall Nozick's original discussion:

> Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that

© Cambridge University Press 2015
Utilitas Vol. 28, No. 3, September 2016
doi:10.1017/S0953820815000424

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life . . . ? . . . Would you plug in? *What else can matter to us, other than how our lives feel from the inside?*[1]

Nozick implies that we shouldn't plug in. He also implies that we wouldn't plug in, and that this shows that we care about things besides our experiences. Nowhere in this discussion does he say anything about welfare. His point is merely that '[w]e learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it'.[2] Nevertheless, for many years, the experience machine was widely regarded as yielding a refutation of hedonism about welfare.

Recently, a growing number of philosophers have challenged this consensus. Some have merely argued that the experience machine doesn't *decisively* rule out hedonism.[3] But others have gone further: although their stated aim has usually been to show that hedonism isn't refuted by the experience machine, they have actually claimed or implied that the experience machine gives us no *pro tanto* reasons to reject hedonism. In their view, the experience machine doesn't rule out hedonism because it doesn't count against it at all.

Harriet Baber explicitly endorses this view. She writes:

Nozick's experience machine thought experiment is generally taken to make a compelling, if not conclusive, case against philosophical hedonism. I argue that it does not and, indeed, that regardless of the results, it cannot provide any reason to accept or reject either hedonism or any other philosophical account of wellbeing . . .[4]

Matthew Silverstein agrees:

Surely, defenders [of the experience machine argument] maintain, our overwhelming intuitive prejudice against the experience machine tells us *something* about the role of happiness in a good life . . . When we contemplate the experience machine scenario, our intuitions tell us emphatically that life on the machine is not the best life for us to live. Surely, these philosophers suppose, our intuitions are at the very least a strike against hedonism. They

[1] Robert Nozick, *Anarchy, State, and Utopia* (New York, 1974), pp. 42–3.

[2] Nozick, *Anarchy, State, and Utopia*, p. 44.

[3] Roger Crisp, *Reasons and the Good* (Oxford, 2006), ch. 4; Katarzyna de Lazari-Radek and Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford, 2014), ch. 9.

[4] Harriet Baber, 'The Experience Machine Deconstructed', *Philosophy in the Contemporary World* 15 (2008), pp. 133–8, at 133.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

must provide some sort of information about our well-being. I have offered three arguments against this appeal to our intuitions.[5]

'We miss the mark', he concludes, 'if we take our intuitions about the experience machine as evidence against hedonism'.[6] Thus, in his view, the experience machine doesn't even give us *pro tanto* reasons to reject hedonism: it is not even 'a strike against hedonism'.

Jason Kawall also contends that the experience machine leaves hedonism unscathed. He writes:

> [T]here are crucial flaws in Nozick's example when taken as an instrument to throw mental state theories into doubt ... Many authors seem to believe that the experience machine example poses a difficulty for mental state accounts of well-being in particular. However, the example does not actually focus on a flaw in mental state theories ... [T]he experience machine example does not pose a difficulty unique to mental state accounts of well-being ...[7]

Kawall's claim is not merely that the experience machine isn't a knockdown refutation of mental state theories (of which hedonism is an example), but that it poses no more of a problem for such theories than it does for other theories: it fails to even 'throw mental state theories into doubt'.

These sentiments are echoed by Sharon Hewitt. Although she initially claims that she will be arguing that the experience machine doesn't yield a 'decisive refutation'[8] of hedonism, she ends up making the stronger claim that it provides no evidence against hedonism:

> Our distaste for living our entire lives hooked up to an experience machine has been cited by Nozick and many philosophers after him as evidence that hedonism is false ... There are at least two basic problems with taking our feelings about the experience machine as evidence that hedonism is false.[9]

In other words, Hewitt objects not merely to the claim that the experience machine gives us *sufficient* reason to reject hedonism, but also to the weaker claim that it gives us *some* reason to do so.

Other philosophers have stopped short of explicitly claiming that the experience machine gives us no evidence against hedonism, but have nonetheless implied or suggested that this is the case. For example, Torbjörn Tännsjö claims that the experience machine argument is 'in an almost obvious way, unsound', and that '[e]ven the validity of

[5] Matthew Silverstein, 'In Defense of Happiness: A Defense of the Experience Machine', *Social Theory and Practice* 26 (2000), pp. 279–300, at 289–90, 299.

[6] Silverstein, 'In Defense of Happiness', p. 296.

[7] Jason Kawall, 'The Experience Machine and Mental State Theories of Well-Being', *Journal of Value Inquiry* 33 (1999), pp. 381–87, at 382–3.

[8] Sharon Hewitt, 'What Do Our Intuitions about the Experience Machine Really Tell Us about Hedonism?', *Philosophical Studies* 151 (2010), pp. 331–49, at 333.

[9] Hewitt, 'Intuitions about the Experience Machine', pp. 337–8.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

the argument can be questioned'. He then claims that 'one more way of understanding Nozick's thought experiment remains', and that on this reading of the thought experiment, the putatively anti-hedonistic intuitions are 'even if true, irrelevant'.[10] One is left with the impression that he takes himself to have shown that every way the experience machine might be invoked against hedonism is a failure.

Alex Barber argues at length that 'the argument Nozick generates from his example is riddled with difficulties'.[11] Although this leaves open the possibility that the experience machine might be the basis for a different, more effective argument against hedonism, Barber doesn't mention this possibility. Instead, he gives a new argument against hedonism that 'takes its cue from Nozick's discussion' but is entirely unrelated to the experience machine.[12] One therefore gets the impression that he believes that the experience machine cannot be effectively employed to cast doubt on hedonism – an impression that is supported by his claim that '[his] quarrel has been with Nozick's *example*, not with what he tried to show with it'.[13]

Finally, although Dan Weijers doesn't wholeheartedly endorse the view that the experience machine gives us no evidence against hedonism, he appears to be tempted by it. For although he says that reflection on the experience machine 'produces some judgments that constitute evidence against hedonism', he claims on the very same page that 'contemplation of the experience machine should no longer give us prima facie reason for rejecting hedonism and other internalist mental state theories of well-being'.[14]

Why have so many philosophers dismissed the experience machine? I believe that this trend is largely attributable to the fact that, almost without exception, discussions of the experience machine have focused on anti-hedonistic arguments that could plausibly be attributed to Nozick – arguments premised on what we *would* do if we were presented with the machine, what we *should* do in such a scenario, or what we *desire* or *value*. These arguments are highly problematic, as many have observed. If alternative ways of invoking the experience machine against hedonism aren't noticed, then it might seem that the experience machine leaves hedonism unscathed.

[10] Torbjörn Tännsjö, 'Narrow Hedonism', *The Journal of Happiness Studies* 8 (2007), pp. 79–98, at 93–5.

[11] Alex Barber, 'Hedonism and the Experience Machine', *Philosophical Papers* 40 (2011), pp. 257–78, at 259.

[12] Barber, 'Hedonism and the Experience Machine', pp. 271–2.

[13] Barber, 'Hedonism and the Experience Machine', p. 276 (italics added).

[14] Dan Weijers, 'Nozick's Experience Machine Is Dead, Long Live the Experience Machine!', *Philosophical Psychology* 27 (2014), pp. 513–35, at 529.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

Many commentators assume that if the experience machine threatens hedonism, it does so because it reveals that we intrinsically desire or value things besides pleasure, or more generally, how our lives feel from the inside. Kawall, Silverstein, Tännsjö, Baber, and Hewitt have all argued that unless we assume that some value- or desire-based theory of welfare is true, this putative fact doesn't imply that pleasure is not the only basic prudential good.[15] Moreover, Kawall, Weijers, and Felipe De Brigard have argued that it is unclear that we do in fact desire or value things besides pleasant experiences: perhaps our unwillingness to permanently to plug into the machine is merely due to *status quo bias* – an irrational preference for maintaining the status quo.[16] After all, you might refuse to *unplug* from the machine if you discovered that your entire life had been spent inside it, and you might refuse to abandon your current life in favour of a more pleasant one even if the latter were merely to take place in a different part of the (real) world.[17]

Some commentators assume that any anti-hedonistic argument based on the machine must proceed via the premise that we *wouldn't* permanently plug in.[18] Hewitt has argued that our resistance to plugging in might be due to practical worries about the machine (e.g. can we be sure that it won't encounter technical difficulties?), and that this possibility cannot be ruled out by stipulating that no such practical problems will arise.[19] Barber has read Nozick as making the obviously problematic argument that (i) since we wouldn't plug in, we aren't hedonists, and that (ii) since we aren't hedonists, hedonism is false.[20] Moreover, Fred Feldman has observed that from the premise that we would refuse to plug into the machine wholly out of concern for our own welfare, it doesn't follow that hedonism is false: for even if hedonism were true, we might refuse to plug in provided that we *believed* it to be false.[21]

[15] Kawall, 'Mental State Theories', p. 385; Silverstein, 'In Defense of Happiness', pp. 285–9; Tännsjö, 'Narrow Hedonism', pp. 94–5; Baber, 'Experience Machine Deconstructed', p. 135; Hewitt, 'Intuitions about the Experience Machine', pp. 332–3, 344.

[16] Kawall, 'Mental State Theories', pp. 382–3; Felipe De Brigard, 'If You Like It, Does It Matter If It's Real?', *Philosophical Psychology* 23 (2010), pp. 43–57, at 50–1; Weijers, 'Nozick's Experience Machine Is Dead', pp. 516–19.

[17] De Brigard, 'If You Like It', pp. 44, 46–50; Kawall, 'Mental State Theories', pp. 382–3.

[18] There is overlap between this group of commentators and the previous one. For many believe that the role of the claim that we wouldn't plug in is to show that we value something besides how our lives feel from the inside.

[19] Hewitt, 'Intuitions about the Experience Machine', pp. 338–41.

[20] Barber, 'Hedonism and the Experience Machine', p. 262.

[21] Fred Feldman, 'What We Learn from the Experience Machine', *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, ed. Ralf M. Bader and John Meadowcroft (Cambridge, 2011), pp. 59–86, at 67–70.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

Some have assumed that if the experience machine counts against hedonism, it does so via the premise that we *shouldn't* permanently plug into it. But from the claim that we shouldn't plug in, nothing follows about welfare. As many have noted, it could be that we shouldn't plug in even though we would be better off if we did: maybe a life spent inside the machine would be undignified or would not befit a human being, and we have strong reasons to live a dignified life. Or perhaps we simply have strong reasons to keep commitments that we couldn't keep if we plugged in.[22]

Those who dismiss the experience machine evidently assume that it cannot be the basis for any plausible anti-hedonistic argument besides the ones just considered. For their dismissal of it is based simply on the fact that those arguments are failures. If they believed that there were any other plausible ways to invoke the experience machine against hedonism, they would have acknowledged this and represented themselves as merely arguing against certain *specific* anti-hedonistic arguments based on the experience machine. But as I will now argue, they are mistaken: there is a better way to use the experience machine against hedonism.

THE EXPERIENTIALLY IDENTICAL LIFETIME COMPARISON ARGUMENT

The arguments considered above appealed to intuitions about what we desire or value, what we would do in certain scenarios, and what we should do in those scenarios. But we also have intuitions that are directly about welfare. Imagine a life of unrelenting physical and psychological torture, in which you achieve nothing, are loved by no one, and hate every moment of your existence. Compare this life with your actual life. To see which of these lives is higher in welfare, you needn't reason from a premise about what you value, or a premise about which life you would, or should, prefer: it is intuitively clear that your actual life is much higher in welfare. Now compare two lives that are identical, except that in one of them, the person takes more pleasure in everything he enjoys. Most people will have the intuition that the life with more pleasure is higher in welfare.

If we want to invoke the experience machine against hedonism, we should consider a case that appeals directly to our intuitions about welfare – not to intuitions about things that are imperfectly correlated with welfare. Not just any such case will do, however: it should be one

[22] Silverstein, 'In Defense of Happiness', pp. 290–3; Chris Heathwood, 'Desire Satisfactionism and Hedonism', *Philosophical Studies* 128 (2006), pp. 539–63, at 553; Ben Bradley, *Well-Being and Death* (Oxford, 2009), p. 10.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

in which we compare two lives (or intervals) that are experientially identical, and thus indistinguishable with respect to the pleasures and pains they contain.

To see why, consider Weijers's Stranger No Status Quo (NSQ) case, which he claims to be the best case involving the experience machine for testing the truth of hedonism.[23] In this case, a stranger has, unbeknownst to him, been 'regularly switched' between real life and the machine, with 50 per cent of his life so far spent in the real world and the other 50 per cent spent in the machine. (This detail is designed to minimize status quo bias.) We stipulate that experiences are 'vastly more enjoyable and varied in the machine', and that 'nearly all of [his] most enjoyable experiences occurred while he was in an Experience Machine and nearly all of his least enjoyable experiences occurred while he was in reality'. The stranger now has to decide whether to plug permanently into the machine or to spend the rest of his life in reality. Which choice would be best *for him*? Weijers polled 77 people and found that 42 of them (55 per cent) thought that it would be best for the stranger to plug in permanently. He concludes that 'experience machine scenarios do *not* provide evidence of widespread agreement about the relative intrinsic prudential value of reality and how our experiences feel to us on the inside', and that 'contemplation of the experience machine should no longer give us a prima facie reason for rejecting hedonism'.[24] The results do not support this interpretation, however. Given that the rest of the stranger's life would be tremendously more enjoyable if he plugged into the machine than if he didn't, it could be that he would be higher in welfare if he plugged in *even though hedonism is false*. For suppose that hedonism is false because pleasure and contact with reality are *both* basic goods. Perhaps the stranger would nevertheless be better off plugging in because the enormous quantities of extra pleasure that he would thereby obtain would be much *better* for him than the contact with reality that he would thereby forego. Thus, if a given life spent in the machine is hedonically better than a given life spent in the real world, you can coherently think that the former life is higher in welfare even though you reject hedonism. For this reason, it could be that hedonism is rejected by many of the subjects who judged that it would be best for the stranger to plug in.

To put hedonism to the test, we should consider a case in which two lives are experientially and thus hedonically identical instead of

[23] Weijers, 'Nozick's Experience Machine Is Dead', p. 528: 'the Stranger NSQ scenario should be used, instead of any of the extant experience machine scenarios, for investigating the intrinsic prudential value of reality and how our experiences feel to us on the inside'.

[24] Weijers, 'Nozick's Experience Machine Is Dead', pp. 525–6, 528–9.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

a case in which the experience machine life is hedonically superior. Consider two lives, *A* and *B*, that are experientially identical and thus identical with respect to the qualitative features, durations, and temporal distribution of the pleasures and pains they contain. The subject of *A* (call him Adam) spends his life in the real world, whereas the subject of *B* (call him Bill) is plugged into an experience machine for his entire life. *A* is a good life of the sort available to citizens of Western countries. Let us stipulate that at no point does Bill interact with, or receive any care from, other human beings: thus, the experience machine runs entirely on its own, without any human intervention. Indeed, at no point after Bill's birth is any person even aware of his existence. (His mother died during childbirth, she alone was aware of the pregnancy, and a robot plugged him into the machine immediately after he was born.)

Do *A* and *B* contain the same total amount of welfare? I believe that many would join me in having the *comparison intuition*: *A* is at least somewhat higher in total welfare than *B*. This is an intuition, not a conclusion derived from other claims. But it is corroborated by the fact that there appears to be something especially *pitiful* about Bill. To pity someone is to feel bad for him on account of the fact that he is in certain respects badly off. Thus, if Bill *warrants* more pity than Adam, he must be lower in welfare.

Some might wonder how those who appear to have the comparison intuition can be sure that they really have this intuition, as opposed to merely having an intuition that *A* and *B* differ along some scale of evaluation *other* than welfare. Perhaps the content of the intuition is that *A* is better 'from the point of view of the universe' than *B*, or that *A* is a more dignified life than *B*. Of course, *certainty* on these matters would not be warranted. But many would take themselves to have the comparison intuition even if they attended carefully to the possibility that *A* and *B* might differ on some other scale. It would be unreasonable to doubt that their intuition is about welfare, just as it would be unreasonable to have analogous doubts about those who claim to have the intuition that if two lives are exactly the same except that the second contains more pleasure, the second life is higher in welfare.

The comparison intuition yields a simple argument against hedonism. A particular thing is *basically* good (bad) for you just if it is good (bad) for you, but not solely in virtue of being suitably related to something else that is good (bad) for you.[25] It is analytic that your welfare during an interval of time is fixed by how basically good or

[25] Others speak instead of basic *intrinsic* goodness (badness), but this creates the misleading impression that it is analytic that how basically good a particular thing is supervenes on its intrinsic features.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

bad for you during that interval everything is.[26] Thus, your *lifetime* welfare is fixed by how basically good or bad for you during your life everything is.[27] Hedonism says that pleasure is the only basic good and pain the only basic bad, and that the *basic prudential value* of a particular pleasure or pain – i.e. *how* basically good (bad) it is – is fixed by its qualitative features (e.g. intensity), its duration, and perhaps its temporal relations to other pleasures and pains.[28] Since *A* and *B* are experientially identical, hedonism implies that they are equal in welfare. Thus, the comparison intuition is incompatible with hedonism. Call this the *experientially identical lifetime comparison argument* – or the *comparison argument*, for short. We can represent it as follows:

(1) *A* is higher in welfare than *B*.
(2) If hedonism is true, then *A* is not higher in welfare than *B*.
(3) *Therefore:* Hedonism is false.

The objections that I considered earlier do not apply to this argument. The argument doesn't involve anyone's having the option of plugging into the machine, and it doesn't rely on any claims about whether we would or should plug in. It is not susceptible to the status quo bias, since the status quo for Adam is the real world and the status quo for Bill is the experience machine. Nor does it rely on any claims about what we desire or value.

The comparison argument exemplifies a familiar strategy: to argue against a view according to which the A-facts supervene on the B-facts, produce a case in which there is intuitively a difference in the A-facts without a difference in the B-facts. Surprisingly, however, the argument has been almost entirely ignored in the literature. To my

[26] This is compatible with the controversial view that if you currently have a future-oriented desire whose object obtains in the future, that object can be basically good for you now. (See Donald Bruckner, 'Present Desire Satisfaction and Past Well-Being', *Australasian Journal of Philosophy* 91 (2013), pp. 15–29, and Dale Dorsey, 'Desire-satisfaction and Welfare as Temporal', *Ethical Theory and Moral Practice* 16 (2013), pp. 151–71.) For even though the time at which the object *obtains* is in the future, the time at which it is *basically good* is now.

[27] We can leave open the possibility of posthumous benefits and harms by stipulating that, by a person's total *lifetime* well-being, we mean the total amount of welfare that she accrues *during her life*. This amount is wholly determined by how basically good or bad for her *during her life* all of the things that are good or bad for her are.

[28] Feldman proposes hedonistic views on which the basic prudential value of a pleasure is determined in part by other factors, such as the truth value of its object. Fred Feldman, *Pleasure and the Good Life* (Oxford, 2004). But as others have argued, it is doubtful that such views are genuine versions of hedonism: see Serena Olsaretti, 'The Limits of Hedonism: Feldman on the Value of Attitudinal Pleasure', *Philosophical Studies* 136 (2007), pp. 409–15, and Michael J. Zimmerman, 'Feldman on the Nature and Value of Pleasure', *Philosophical Studies* 136 (2007), pp. 425–37. I explain how a variant on the present argument can be used to argue against such views in Eden Lin, 'Pluralism about Well-Being', *Philosophical Perspectives* 28 (2014), pp. 127–54, at 140–5, 146–7.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

knowledge, only Roger Crisp discusses it – and very briefly.[29] The argument is not mentioned by any of the philosophers who claim that the experience machine leaves hedonism unscathed – even those whose articles postdate Crisp's discussion. Indeed, in a 2013 survey article purporting to be 'exhaustive of all the published paper-length attempts to refute the experience machine objection to hedonism during the last 15 years and indicative of all of the types of critical responses to the experience machine objection to hedonism to date', Dan Weijers and Vanessa Schouten make no mention of anything like the comparison argument.[30] Instead, they represent every anti-hedonistic argument involving the experience machine as relying on the claims (i) that a life in the experience machine would be experientially *superior* to a life in the real world, and (ii) that we *value* contact with reality or *prefer* life in the real world to life in the machine.[31]

The comparison argument is valid, and its second premise is unquestionably true, but it isn't a knockdown refutation of hedonism. For however intuitively compelling its first premise is, it is not indubitable. Since I doubt that the experience machine could be the basis of an even more decisive argument against hedonism, I agree with those who claim that it doesn't yield a knockdown refutation of hedonism. On the other hand, I believe that the comparison argument gives us *pro tanto* reasons to reject hedonism. In other words, the argument is some evidence for the falsity of hedonism. When tallying up the pros and cons of hedonism, the fact that it is incompatible with the comparison intuition should be listed as one of the cons.

This might seem obvious. If a theory has a counterintuitive implication, then presumably this counts against the theory at least to some degree. So surely, hedonism's incompatibility with the comparison intuition is a *pro tanto* reason to reject hedonism. But there are two reasons why some might disagree. First, the comparison argument might be suspected of begging the question. Second, it might be argued we can undercut anti-hedonistic intuitions by telling a plausible debunking story about their origins. In the remainder of the article, I

[29] Crisp, *Reasons and the Good*, p. 118. Heathwood comes close: he discusses an argument in which it is claimed that, if you were 'just taking yourself into account', you would *prefer* a life in the real world to an experientially identical life in an experience machine. Chris Heathwood, 'Welfare', *The Routledge Companion to Ethics*, ed. John Skorupski (London, 2010), pp. 645–55, at 649–50. Feldman, 'What We Learn', pp. 80–1, considers a lifetime comparison argument that appeals directly to intuitions about welfare, but in which the life in the experience machine is hedonically *superior*.

[30] Dan Weijers and Vanessa Schouten, 'An Assessment of Recent Responses to the Experience Machine Objection to Hedonism', *Journal of Value Inquiry* 47 (2013), pp. 461–82, at 469.

[31] Weijers and Schouten, 'Recent Responses to the Experience Machine', pp. 466–7.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

will argue that these are not good reasons to deny that the comparison argument counts against hedonism.

BEGGING THE QUESTION

Some might worry that the comparison argument begs the question against hedonism. After all, since hedonism straightforwardly entails that A and B are equal in welfare, any committed hedonist would surely respond to the argument by rejecting its first premise. And it might be thought that if an argument employs a premise that would surely be rejected by anyone who is committed to the negation of its conclusion, then that argument begs the question. This thought strikes me as mistaken, however. For it implies that *every* obviously valid single-premise argument begs the question. After all, a *committed* believer in a given claim is someone who would reject anything she believes to be incompatible with that claim. And for any *obviously* valid single-premise argument whose conclusion is p, any committed believer in $\sim p$ would recognize that the premise of the argument is incompatible with $\sim p$. Thus, every obviously valid single-premise argument employs a premise that would surely be rejected by anyone who is committed to the negation of the argument's conclusion. But presumably, not every such argument begs the question: it wouldn't be question-begging to argue that since Socrates is human but not female, not all humans are female.[32]

More importantly, even if the aforementioned condition is sufficient for begging the question in *some* sense, this doesn't matter. I don't need to show that the comparison argument doesn't beg the question in *any* sense. What's at issue is whether the experience machine provides evidence for the falsity of hedonism. On views like the one just discussed, begging the question is a purely dialectical vice, not an epistemic one. Even if the comparison argument begs the question according to such views, it doesn't follow that it gives us no evidence for the falsity of hedonism.[33] Although any committed hedonist would reject the claim that A is higher in welfare than B, many people will have the intuition that A is higher in welfare than B, and those people

[32] Moreover, notice that the comparison argument cannot be thought to beg the question simply because its premises cannot all be true unless its conclusion is true: for *all* valid arguments are like this, but not all valid arguments beg the question. Nor can it be thought to beg the question simply because any reason to doubt its conclusion would also be a reason to doubt that all of its premises are true: once again, *all* valid arguments are like this. These familiar points are made in Frank Jackson, *Conditionals* (Oxford, 1987), pp. 100–1.

[33] See James Pryor, 'What's Wrong with Moore's Argument?', *Philosophical Issues* 14 (2004), pp. 349–78. Pryor argues that an argument can provide justification for its conclusion even if it is dialectically ineffective relative to certain audiences.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

will therefore have some evidence that hedonism is false. Indeed, the committed hedonist could be among those who possess this evidence, since she could share the relevant intuition in spite of her commitment to hedonism.

Similar remarks apply to Alan Hazlett's account of begging the question, according to which a use of an argument begs the question just if at least one of its premises is such that the utterer of the argument cannot reasonably believe that it is accepted (i.e. assumed) in the conversation.[34] Suppose I am talking to someone I know to be a committed hedonist. I know that if I were to utter the comparison argument, she would reject its first premise. Thus, I cannot reasonably believe that the first premise of the argument is accepted in this conversation. On Hazlett's view, this shows that the comparison argument begs the question in this context. But it doesn't follow that the argument provides us, or the hedonist, with no evidence that hedonism is false. For we (and the hedonist) could have the intuition that *A* is higher in welfare than *B* even though I cannot reasonably believe that this proposition is accepted in the present conversation.

In order for my thesis to be undermined, it would have to be shown that the comparison argument begs the question in a way that renders it evidentially inert – i.e. a way in virtue of which it gives us no *pro tanto* reasons to reject hedonism. I cannot consider every way one might attempt to do this, but let me address two possibilities.

According to Frank Jackson, an argument begs the question just if at least one of its premises is such that any sane person who doubts its conclusion would have background beliefs relative to which the evidence for that premise is no evidence for it. (Consider the argument, 'Very many experts believe that others have minds; it is sensible to believe what very many experts believe; therefore, it is sensible to believe that others have minds.' On Jackson's view, this begs the question because the evidence for the first premise must be behavioural, and anyone sane who doubts the conclusion of the argument would doubt that behaviour can be evidence for mental states.)[35] If the comparison argument begged the question on Jackson's account, it would arguably provide no evidence for the falsity of hedonism (at least to people who don't already believe that hedonism is false). But the argument doesn't beg the question on this account. For the only premise in virtue of which the argument *might* beg the question is the first premise: that *A* is higher in welfare than *B*. (The second premise – that *A* is not higher in welfare than *B* if hedonism is true – is clearly

[34] Alan Hazlett, 'Epistemic Conceptions of Begging the Question', *Erkenntnis* 65 (2006), pp. 343–63, at 354–8.

[35] Jackson, *Conditionals*, pp. 110–12.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

true, and is common knowledge.)[36] The evidence for the first premise is provided by an evaluative intuition. But it isn't true that anyone sane who doesn't believe that hedonism is false would have background beliefs relative to which an evaluative intuition cannot provide evidence for that premise. Indeed, even a committed hedonist could grant that evaluative intuitions provide evidence.

On Walter Sinnott-Armstrong's view, a particular use of an argument begs the question just if at least one of its premises is such that any reason the audience has to believe that premise depends either on (i) their already believing the conclusion or (ii) their already having some reason to believe the conclusion.[37] If the comparison argument begged the question on this account, then it could not provide evidence for the falsity of hedonism to anyone who didn't either already believe that hedonism is false or already have some reason to believe that it's false. But the argument does not beg the question on this account, regardless of whether its audience believes hedonism, disbelieves it, or is agnostic about it. For someone could have an intuition that *A* is higher in welfare than *B* (and thus a reason to believe that *A* is higher in welfare than *B*), even though he doesn't already believe that hedonism is false. And someone could have such an intuition even though he doesn't already have some reason to believe that hedonism is false. For example, a hedonist who doesn't already have any reason to doubt the truth of hedonism (and who doesn't believe, inconsistently, that hedonism is false) could acquire a reason to believe that *A* is higher in welfare than *B* by having an intuition to this effect.

I suspect that those who are tempted to regard the comparison argument as question-begging actually have the following in mind. Perhaps most hedonists have already considered pairs of lives like *A* and *B*, and they have already come to the conclusion that, even though it might intuitively *seem* that such lives can differ in welfare, they cannot. If so, then most hedonists accept hedonism in full awareness of the fact that it contradicts the comparison intuition. The comparison intuition will therefore do nothing further to reduce

[36] Again, I am assuming an understanding of hedonism on which it straightforwardly entails the second premise. See n. 28.

[37] Walter Sinnott-Armstrong, 'Begging the Question', *Australasian Journal of Philosophy* 77 (1999), pp. 174–91, at 182–3. I am focusing here on uses of arguments for the purpose of *audience justification*, i.e. trying to show the audience that they have a reason to believe the conclusion. Sinnott-Armstrong says that when the purpose is *arguer justification* (i.e. showing that the arguer has a reason to believe the conclusion), what matters is whether the *arguer* would have any reason to believe the premise if he didn't already believe the conclusion and didn't already have some reason to believe the conclusion. What I say below could straightforwardly be adapted to show that the comparison argument wouldn't beg the question even if it were used for arguer justification.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

their credences in hedonism: it has already been 'priced in' to those credences.

Even if those claims are true, however, they don't cast doubt on the fact that the comparison argument provides us with *pro tanto* reasons to reject hedonism. Whether something is a *pro tanto* reason to reject a theory doesn't depend on whether adherents of the theory have already considered it. It could be that most hedonists accept hedonism in full awareness that its incompatibility with the comparison intuition is a *pro tanto* reason to reject it: it's just that they believe this reason to be outweighed by reasons to accept it. Indeed, the thought that the intuition has already been 'priced in' to their credences suggests that they may have already lowered those credences in response to it. Such an adjustment would be warranted only if the intuition provides evidence against hedonism. Moreover, it is not obvious that most hedonists are already aware that hedonism is incompatible with the comparison intuition – though they would presumably become aware of this as soon as they considered the issue. As I said earlier, there has been very little discussion of the comparison argument in the literature.

As far as I can see, the comparison argument doesn't beg the question in any way that suggests that it fails to give us *pro tanto* reasons to reject hedonism.

DEBUNKING ANTI-HEDONISTIC INTUITIONS

There is another reason one might doubt that the comparison argument provides evidence against hedonism: perhaps we can undercut anti-hedonistic intuitions by telling a debunking story about their origins. This has been suggested by Silverstein, and more briefly, by Crisp and Hewitt.[38] I will focus on rebutting Silverstein's version of this objection, since it is the most detailed one. But the core of my response could be adapted to Crisp and Hewitt.

Drawing on some remarks by Brandt and Railton, Silverstein gives the following argument for dismissing anti-hedonistic intuitions (such as the comparison intuition). When acting on a desire leads to pleasure, this desire is positively reinforced in the sense that we are more likely to retain it and to act on it in the future. When acting on a desire leads to displeasure, this desire is negatively reinforced. Thus, the

[38] Silverstein, 'In Defense of Happiness', pp. 293–8; Crisp, *Reasons and the Good*, pp. 120–2; and Hewitt, 'Intuitions about the Experience Machine', pp. 343–7. Only Silverstein makes the strong claim that such an argument can show that anti-hedonistic intuitions have no evidential force. Arguments of this kind are also discussed in David Sobel, 'Varieties of Hedonism', *Journal of Social Philosophy* 33 (2002), pp. 240–56, at 244–7.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

desires that we have and act on are largely ones that have led to pleasure when we have acted on them in the past. Now, pleasure is more easily achieved by someone who intrinsically desires other things than by someone who intrinsically desires only pleasure: this is the *paradox of hedonism*. Thus, we should expect to have intrinsic desires for things besides pleasure, since these desires will have been positively reinforced via their tendency to produce pleasure. And we should also expect not to have a desire to maximize pleasure at any cost, since such a desire will have been negatively reinforced. In this way, pleasure ultimately explains why we have all of the desires that we have – including our intrinsic desires for things other than pleasure. Moreover, our evaluative intuitions reflect our intrinsic desires, in the sense that the things that intuitively strike us as basically good for us tend to be precisely the ones that we intrinsically desire.[39] Silverstein (who uses the term 'happiness' interchangeably with 'pleasure') claims that this shows that we should distrust our anti-hedonistic intuitions:

> [O]ur experience machine intuitions reflect our desire to remain connected to the real world, to track reality. But . . . the desire to track reality owes its hold upon us to the role it has played in the creation of happiness . . . Our intuitive views about what is prudentially good, the views upon which the experience machine argument relies, owe their existence to happiness. We miss the mark, then, if we take our intuitions about the experience machine as evidence against hedonism . . . Even though it leads us away from happiness in the case of the experience machine, our desire to track reality points indirectly to happiness . . . The mere existence of our intuitions against the experience machine should not lead us to reject hedonism. Contrary to appearances, those intuitions point – albeit circuitously – to happiness. And as a result, they no longer seem to contradict the claim that happiness is the only thing of intrinsic prudential value.[40]

Because our intuitions about basic goodness reflect our intrinsic desires, and because we have whatever intrinsic desires have led to pleasure when acted upon, we should not regard our anti-hedonistic intuitions as evidence against hedonism. Thus, Silverstein concludes, we should reject even the weak claim that our anti-hedonistic intuitions about the experience machine 'are at the very least a strike against hedonism' and that they at least 'provide some sort of information about our well-being' (i.e. some evidence that hedonism is false).[41]

Before I evaluate this argument, let me explain what it would take for it to succeed. Whether hedonism is true is a question in the normative ethics of welfare, and debates in the normative ethics of welfare cannot get started without evaluative intuitions. One cannot

[39] Silverstein, 'In Defense of Happiness', pp. 293–6.
[40] Silverstein, 'In Defense of Happiness', p. 296.
[41] Silverstein, 'In Defense of Happiness', p. 299.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

even begin to make a case for hedonism without appealing to intuitions about pleasure and pain (e.g. that if two lives are identical, except that one of them contains more pleasure than the other, the life containing more pleasure is higher in welfare). Analogous claims are true of any other theory of welfare. Thus, the project that we and Silverstein are engaged in assumes that at least *some* evaluative intuitions are evidence. If Silverstein's argument turned out to show that none of our evaluative intuitions provide evidence for anything, then it would show too much. Thus, the aim of his argument is not to undermine the evidential force of *all* evaluative intuitions, but to show that *anti-hedonistic* intuitions (such as the comparison intuition) are *especially* unreliable. If the argument discredited all of our intuitions equally, it would undermine the entire normative ethical debate about welfare, rather than constituting a gambit on behalf of hedonism in that debate. Silverstein is suggesting that because *pleasure* is the thing that ultimately explains why we have the intuitions that we do, those among our intuitions that count against hedonism are – *unlike other intuitions* – devoid of evidential force.

Silverstein's argument does *not* discredit anti-hedonistic intuitions more than it does other intuitions, however. The conclusion that our anti-hedonistic intuitions should be distrusted is meant to follow from two crucial claims: (i) that our evaluative intuitions reflect our intrinsic desires and (ii) that our intrinsic desires are formed in such a way that we have no reason to expect their objects to be basically good for us. But if those two claims imply that our anti-hedonistic intuitions are unreliable, they imply the same thing about *all* of our evaluative intuitions. After all, Silverstein claims that *all* of our intuitions (not just the anti-hedonistic ones) reflect our intrinsic desires, and that *all* of our intrinsic desires (not just the ones for things besides pleasure) are formed in such a way that we have no reason to expect their objects to be basically good for us. The fact that, on his picture, it is *pleasure* that ultimately explains our desires may create the impression that anti-hedonistic intuitions are *especially* unreliable, or that hedonistic intuitions are immune to the argument. But this impression is mistaken. The story about how pleasure explains our desires plays no role besides supporting the claim that, given how our intrinsic desires are formed, there is no reason to expect their objects to be basically good for us. The argument would be no worse off if Silverstein had suggested instead that all of our desires are reinforced in accordance with whether they lead to food, sex, or something else. As long as we accept *some* causal story about the formation of our desires on which we have no reason to think that we intrinsically desire things in proportion to how basically good they are, and as long as we grant that the things that intuitively strike us as basically good tend

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

to be precisely the things that we intrinsically desire, the argument condemns all of our evaluative intuitions equally.[42]

Silverstein might dispute this. He claims that his story about the reinforcement of desires not only shows that anti-hedonistic intuitions are unreliable but also suggests that hedonism is *true*. This cannot be the case if the story's only role is to support the claim that our intrinsic desires are formed in such a way that we have no reason to expect their objects to be basically good for us. But an examination of what he says about this strengthens rather than weakens my case. He says that his story is evidence for hedonism because

> Hedonism explains what would otherwise appear to be a mysterious coincidence, namely, the fact that all of our desires point towards happiness . . . [W]ithout hedonism we cannot explain *why* all of our desires are related to happiness in this way. The most plausible explanation is a hedonistic one: the reason all of our desires point towards happiness is that happiness is the only thing that is intrinsically prudentially valuable.[43]

Silverstein does not tell us *how* the truth of hedonism could explain the truth of the psychological story about desire-formation that he proposes. Though I do not object in general to the idea that an evaluative truth could explain a descriptive one, I don't see how this particular explanation would go. Unless we posit a benevolent deity, wouldn't it be a mysterious coincidence that all of our desires are reinforced in accordance with how well they lead to the sole basic good? If we believed that our desires are reinforced in accordance with whether they lead to the propagation of our genes, we wouldn't conclude that the propagation of our genes is the sole basic good – and rightly so.[44] Thus, the impression that Silverstein's story supports hedonism is an illusion.[45] Once we see this, it is even clearer that the story does no work besides supporting the claim that we have no reason to think

[42] See Sharon Street, 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127 (2006), pp. 109–66, at 155. Street claims that a dilemma analogous to her Darwinian dilemma for value realism 'could be constructed using any kind of causal influence on the content of our evaluative judgements'.

[43] Silverstein, 'In Defense of Happiness', p. 297.

[44] This assumes the falsity of the reductive view that to be a basic good *just is* to be something that plays the role vis-à-vis our desires that Silverstein says is played by pleasure. I find this view implausible, but I cannot prove that it is false here. Silverstein and other defenders of hedonism against the experience machine do not endorse such a view.

[45] Silverstein only considers a version of the experience machine argument in which the fact that we desire things besides pleasure is thought to count against hedonism. Perhaps his claim is that any advocate of *this* version of the argument should regard the truth of the psychological story about desire formation as evidence that pleasure is the only basic good, since any such advocate already posits a tight connection between what we desire and what is basically good for us. Whether or not he is right about this, I maintain that the truth of his psychological story would not in fact support hedonism.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

that we intrinsically desire things in proportion to how basically good they are. For even if the role of pleasure in the story were replaced by something else (e.g. the propagation of our genes), the argument would raise no less of a worry about the reliability of our intuitions.

Thus, Silverstein has told a story about the formation and retention of our desires (and about the relationship between our evaluative intuitions and our desires) that threatens to debunk *all* of our evaluative intuitions – including those that support hedonism. The fact (if it is one) that our desires are reinforced in accordance with whether they lead to *pleasure* does not show that hedonistic intuitions are more reliable than other ones. For, as I have just argued, the supposed fact that our desires are reinforced in this manner is not evidence for the truth of hedonism.[46]

With these considerations in view, we can offer two responses. First, Silverstein hasn't given us sufficient reason to believe that our evaluative intuitions reflect our intrinsic desires and that our desires are reinforced in accordance with whether they lead to pleasure. The psychological picture that he sketches is, at best, a *plausible hypothesis*. Thus, even if the truth of this picture *would* imply that anti-hedonistic intuitions are no evidence against hedonism, as things stand, this conclusion would be premature. Since we have insufficient reason to believe it, Silverstein's picture doesn't show that the comparison intuition provides no evidence *at all* against hedonism. (Given the current state of the sciences, the same would likely be true of many other empirical hypotheses about the origins of our evaluative intuitions that might be invoked to debunk anti-hedonistic intuitions.)

Second, and more importantly, Silverstein's argument proves too much. If it shows that our anti-hedonistic intuitions are no evidence against hedonism, then it also shows that all of our evaluative intuitions are no evidence for anything. Perhaps this will turn out to be true at the end of the day: perhaps a debunking argument will be produced that undermines all of our evaluative intuitions. But this is a possibility that threatens to undermine normative ethics entirely, not one that can selectively be invoked in favour of one normative ethical view over another.[47] If, for the time being, we are entitled (as I assume

[46] Similarly, the putative fact that we have the evaluative outlook that we do because it enhances reproductive fitness doesn't show that the judgement that reproductive fitness is good is more reliable than the rest of our evaluative judgements. Street, 'Darwinian Dilemma', argues that the evolutionary influence on our evaluative outlook undermines realism about value altogether; she doesn't claim that it supports a view on which reproductive fitness is the sole, or central, good.

[47] For a similar point about some attempts to use evolutionary debunking arguments against certain views in normative ethics, see Guy Kahane, 'Evolutionary Debunking Arguments', *Noûs* 45 (2011), pp. 103–25.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press

we are) to continue discussing whether hedonism is true, it cannot be claimed that anti-hedonistic intuitions are of no evidential value. If the comparison intuition is true, then hedonism is false. We cannot be sure that the intuition is true, so it doesn't give us a knock-down argument – one that would warrant us in rejecting hedonism *come what may*. Theory choice is a holistic matter, and we cannot just tally up hedonism's disadvantages while ignoring its advantages. Nor can we ignore the possibility that the intuition is misleading. But the fact that hedonism contradicts the intuition *is* a disadvantage – i.e. a *pro tanto* reason to think that hedonism is false.

CONCLUSION

I have argued for a moderate claim: although the experience machine doesn't yield a decisive refutation of hedonism, it gives us some evidence against hedonism. Some readers might think that this claim is obviously true, and thus not worth arguing for. I would refer them to the many recent discussions of the experience machine in which the claim is denied. In arguing for this claim, I don't pretend to have established anything especially surprising. My aim has merely been to correct one of the excesses of the recent literature.

I have also argued that those who wish to use the experience machine to cast doubt on hedonism should not appeal to claims about what we value, what we would do if given the option of plugging into the machine, or what we should do if given this option. Instead, they should appeal to an intuition about the welfare values of two experientially identical lives, one of which is spent inside the experience machine. It is my hope that the comparison argument will become the standard way of invoking the experience machine against hedonism.[48]

eden.lin@rutgers.edu

[48] I thank Gwen Bradford, Richard Yetter Chappell, Fred Feldman, Daniel Greco, Chris Heathwood, Frank Jackson, Guy Kahane, Adam Lerner, Errol Lord, Barry Maguire, Gideon Rosen, Matthew Silverstein, Daniele Sgaravatti, Pablo Stafforini, and an anonymous referee.

https://doi.org/10.1017/S0953820815000424 Published online by Cambridge University Press