
The Experience Machine Deconstructed

Harriet Baber

University of San Diego
babber@sandiego.edu

ABSTRACT: 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 since it presupposes preferentism, the desire-satisfaction account of wellbeing. Preferentists cannot take any comfort from the results of such thought experiments because they assume preferentism and therefore cannot establish it. Neither can anyone else, since only a preferentist should accept the terms of the thought experiment.

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 since it presupposes preferentism, the desire-satisfaction account of wellbeing. Preferentists cannot take any comfort from the results of such thought experiments because they assume preferentism and therefore cannot establish it. Neither can anyone else, since only a preferentist should accept the terms of the thought experiment.

1. The Experience Machine

Fantasies intended to show the folly of hedonism are cheap and plentiful: when we consider hedonically optimal states of affairs, from traditional dystopias like *Brave New World* to *The Matrix* and, on reflection, most of us conclude that life in a fools' paradise, doped, deceived and deluded, isn't worth it.

Nozick's Experience Machine thought experiment is supposed to go one better.

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

In Nozick's thought experiment we are invited to consider the prospect of a fools' paradise from the first person perspective and to choose: would we prefer the perfect hedonic illusion that the experience machine delivers or a hedonically inferior state of (really) doing certain things and being a certain way. Nozick suggests that we should choose the latter:

What does matter to us in addition to our experiences? First, we want to *do* certain things, and not just have the experience of doing them...A second reason for not plugging in is that we want to *be* a certain way, to be a certain sort of person...Thirdly, plugging into an experience machine limits us to a man-made reality (Nozick, 42-43).

According to the terms of the thought experiment, we are to take the preferences of subjects, revealed in their choices, as decisive: if subjects choose to plug into the machine we are to conclude that the hedonically optimal states it delivers are best for them; if they choose to forgo these hedonically optimal states than we should infer that such states are not best for them but rather that there are other factors besides pleasure or happiness, presumably ways of being and doing, that contribute to wellbeing.

The thought experiment is an ambitious piece of conceptual analysis. The aim is not merely to establish that most people would choose reality over life in a fools' paradise. That is consumer research. It is not merely to determine what subjects "value" (prudentially) or what "matters" to them—what they believe, whether rightly or wrongly, is good for them. That is sociology. The purpose of the thought experiment is to elicit subjects informed, rationally considered preferences as revealed in their choices under epistemically favorable conditions because the assumption is that under these favorable conditions most subjects will get it right: the states which matter to them will be the states that really *matter*—those which in fact contribute to wellbeing. The thought experiment thus turns on the following plausible but, as we shall see, problematic assumption:

P: If a reasonable and informed subject, *i*, would choose *S* over *S'*, then *S* would contribute more to *i*'s wellbeing than *S'*.

Most subjects as it turns out prefer hedonically inferior states of doing and being to life on the machine so we are supposed to conclude that hedonism is false.

The choices of subjects to forgo the pleasures of the experience machine do not however undermine philosophical hedonism, the doctrine that what *makes* a state of affairs contribute to wellbeing is it's being pleasurable. Indeed, whatever the results,

the thought experiment cannot either confirm or disconfirm hedonism or any other philosophical account of prudential value.

2. Philosophical Accounts of Wellbeing

To see this it is important to distinguish philosophical accounts of wellbeing, which purport to identify what it is about states of affairs that *make* them contribute to wellbeing, that is to elicit our *criteria* for wellbeing, from empirical conjectures about what sorts of states in fact satisfy such criteria.

Philosophical hedonism is the doctrine that pleasure is criterial for wellbeing so that what *makes* a state of affairs *S* better for an individual *i* than a state *S'* is its being hedonically superior to *S'*. Philosophical preferentism is a competing account according to which what *makes* a state of affairs good for a person is her desiring it, so that a state, *S*, is better for an individual *i* than a state *S'*, if *S* is higher on *i*'s preference-ranking. This is a hard saying for some who imagine that we desire things because we believe them to be good rather than believing them to be good (at least for us) because we desire them, but that is the preferentist criterion for wellbeing. Philosophical preferentists do not hold that preference satisfaction is merely good evidence for wellbeing or that it is an inevitable concomitant of whatever it is that makes for wellbeing, but rather that it is the very thing that *makes* our lives go well.

Preferentism is often taken to be the account favored by orthodox economists. Most mainstream economists however are not philosophical preferentists. Most are either agnostic about what wellbeing is or assume that it is pleasure/happiness and that preference-satisfaction is a good proxy. "In the middle of the nineteenth century," one welfare economics textbook announces, "it was popular in some philosophical circles to assume that pleasure and pain could be numerically measured."

Some of the nineteenth century advocates of utility calculus thought utility could be standardized and measured...But no one has yet succeeded in defining an objective unit of utility...The problem with asking about utility is this. If you ask 'How many units of happiness would you now get if I give you a banana?' you will be laughed at...Ask instead, 'Would you prefer a banana or an apple?'...*The theory of preferences...is connected to, and is a generalization of, the old-fashioned nineteenth century theory of utility* (Feldman, 11-2; emphasis added).

The author of this passage is a good old-fashioned hedonist who believes that, as a matter of empirical fact, we generally prefer what makes us happy and that for theoretical purposes we should operate with preferences because, unlike pleasure or happiness, preference is objective and quantifiable. On this account, choice reveals preference, which is a proxy for wellbeing—whatever it is—so individuals' choices tell us which states of affairs are best for them.

Orthodox economists' faux-preferentism however is not a philosophical account of wellbeing but a methodological assumption. And it is provisional. So good old-

fashioned hedonists like Daniel Kahneman, et. al. in their evocatively titled “Back to Bentham” urge colleagues to abandon faux-preferentism because the results of empirical research show that happiness and preference-satisfaction do not go in tandem and, on the positive side, because they hold that with improved techniques happiness is both observable and measurable. This is exactly the response philosophical hedonists should make to the results of experiments which suggest that subjects do not choose hedonically optimal states—including the results of thought experiments like the Experience Machine.

More generally, anyone who takes preference-satisfaction merely as evidence for whatever it is that constitutes wellbeing should regard P as an empirical generalization and adopt the same policy: if subjects do not choose states that satisfy their criteria for wellbeing, so much the worse for P. Philosophical preferentists by contrast should regard P as a conceptual truth which follows from their understanding of wellbeing as preference satisfaction and the plausible view that what we choose in epistemically favorable circumstances is what we prefer

According to the terms of the Experience Machine thought experiment, the states subjects choose are the states that are best for them. Preferentists should accept these terms but there is no reason why anyone else should. There is no reason why hedonists, perfectionists, or theologians, who hold that what is good for us is pleasing God, should accept these terms, or be troubled by the results of any comparable experiment showing that states of affairs which satisfy their criteria for wellbeing are rejected by all or most subjects. If it turns out that subjects reject options that produce maximal pleasure, perfect human nature or please God they should rather conclude: “More fool they—they just don’t want what’s good for them!”

3. Does The Experience Machine Refute Hedonism?

If this is correct, then when it comes to the philosophical question of what *makes* a state of affairs contribute to wellbeing the Experience Machine thought experiment is uninformative regardless of what most, or even all, subjects choose.

Assume that all subjects choose to get on the machine, presumably the best-case scenario for the hedonist. This shows that there is a correlation between preference-satisfaction and happiness, because as it happens informed subjects choose, and so prefer, hedonically optimal states.

Suppose you are a hedonist. This result is good news to the extent that it shows that preference, revealed in choice, is good evidence for wellbeing. It provides empirical support for P, which you understand as an empirical claim. It does not however show that pleasure or happiness is criterial for wellbeing or that other characteristics are not criterial for wellbeing, and so does not either confirm hedonism or disconfirm competing accounts.

Suppose you are neither a hedonist nor a preferentist but hold that there is some property, X, other than pleasure, happiness or preference-satisfaction, which is criterial for wellbeing. You can dismiss the results of the thought experiment as completely irrelevant: happiness and informed desire-satisfaction go together, certainly, but neither of them is what matters for wellbeing. Whether you hold that

the value of X is personal autonomy, virtue, following the ways of the elders, making as much money as possible, or pleasing God the results of the thought experiment are of no consequence: wellbeing is a matter of being autonomous, being virtuous, following the ways of the elders, making lots of money or pleasing God, even if that doesn't either make us happy or satisfy our preferences.

Suppose you are a preferentist, who holds that preference-satisfaction is criterial for wellbeing and so regards P as a conceptual truth. Then subjects' choices cannot count against P. If all subjects choose hedonically optimal states this shows that the sort of states that satisfy your criterion for wellbeing, states which rational, informed subjects choose and hence prefer, are hedonically optimal states: it does not establish that the hedonic character of such states is what *makes* them contribute to wellbeing. You should conclude that hedonically optimal states are good for people *because* they are what people prefer.

If this is correct then the choices of subjects to plug into the machine show nothing: they do not provide any evidence for hedonism or undermine competing philosophical accounts of wellbeing. Even on the best case scenario, the hedonist cannot win.

However the hedonist cannot lose either.

Assume that all subjects choose not to get on the machine. Neither the hedonist, nor any one else, should worry about this outcome, though none have any cause for celebration either. Preferentists will conclude that, as it happens, the states of affairs that satisfy their criterion for wellbeing are states of real being and doing: these states are good for people *because* people prefer them. Everyone else should regard the result as irrelevant.

Suppose you are a hedonist. You will note that even in epistemically favorable circumstances people do not choose what is good for them and hence that preference-satisfaction is not a suitable proxy for welfare, which you understand as pleasure or happiness, and so reject economists' faux-preferentism. Like Kahneman, et. al., you will reject P—not hedonism.

Suppose you are an advocate of real being and doing. *Prima facie*, this is a cause for rejoicing. But not so fast: you have a dilemma. Would you have dismissed the results of the experiment—and so rejected P—if all subjects had chosen to get on the machine? If so, then the thought experiment is no more than window-dressing: nothing ventured, nothing gained. You regard P as an empirical generalization about people's proclivity for choosing states of real being and doing which, by your lights, are better for them than life in a fools' paradise.

If not, then you are a closet preferentist: you regard P as a conceptual truth. The results of the thought experiment merely show that the states of affairs which satisfy the preferentist criterion for wellbeing are states of real being and doing rather than hedonically superior illusions.

Maybe this is all the thought experiment was supposed to show, viz. that rational, informed agents do not always prefer hedonically optimal states and so that *as preferentists* we should not imagine that such states invariably make people better off. But, if so, then the thought experiment is not, as it is commonly taken to be, a "weighty objection to hedonism of every kind" (Crisp). Since hedonists, being hedonists, are not preferentists they have no reason to take the states of affairs

subjects choose as those that are best for them. If subjects choose hedonically inferior states of real being and doing, more fool they: they don't want what's good for them.

4. We Are All Preferentists Now

Regardless of what subjects choose, the Experience Machine cannot either confirm or disconfirm any philosophical theory of wellbeing. It merely tests the empirical hypothesis that informed choosers prefer hedonically optimal states. Preferentists get nothing from the thought experiment since it presupposes preferentism and so, regardless of the results, cannot provide any further support for it. No one else gets anything out of it since no one else should take subjects' choices as decisive in establishing what is good for them. So the thought experiment is either question begging or irrelevant.

However, even if the nature of choices subjects make cannot support or undermine any philosophical account of wellbeing, the fact that we believe that they do is telling. Without making fine distinctions between what *makes* a state of affairs contribute to wellbeing and which sorts of states have what it takes, we assume that the states informed subjects choose make them better off. And this suggests that we are all preferentists now.

Even if the Experience Machine can't pump intuitions that would help us decide between competing philosophical accounts of wellbeing, there may be another thought experiment that can, viz. the Meta Experience Machine. *Suppose there were a machine that reliably determined what sorts of states subjects would prefer if they were fully informed about all relevant facts. Would such a machine tell us what sorts of states were good for people?* Most of us think it would: that is why we regard the Experience Machine thought experiment as informative. But only preferentists should regard it as informative. Therefore most of us are preferentists.

Works Cited

- Crisp, Roger. 2006. "Well-Being" *The Stanford Encyclopedia of Philosophy*.
<http://plato.stanford.edu/archives/win2006/entries/well-being/>.
- Feldman, Allan M. 1980. *Welfare Economics and Social Choice Theory*. Norell, MA: Kluwer Academic Publishers.
- Daniel Kahneman et al. **CITATION INFO NEEDED**
- Nozick, Robert. 1974. *Anarchy, State, Utopia*. New York: Basic Books.