

The Experience Machine

Robert Nozick

...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? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your *next* two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. Others can also plug in to have the experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everybody plugs in.) Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* Nor should you refrain because of the few moments of distress between the moment you've decided and the moment you're plugged. What's a few

moments of distress compared to a lifetime of bliss (if that's what you choose), and why feel any distress at all if your decision *is* the best one?

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. In the case of certain experiences, it is only because first we want to do the actions that we want the experiences of doing them or thinking we've done them. (But *why* do we want to do the activities rather than merely to experience them?) A second reason for not plugging in is that we want to *be* a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank. Is he courageous, kind, intelligent, witty, loving? It's not merely that it's difficult to tell; there's no way he is. Plugging into the machine is a kind of suicide. It will seem to some, trapped by a picture, that nothing about what we are like can matter except as it gets reflected in our experiences. But should it be surprising that what *we are* is important to us? Why should we be concerned only with how our time is filled, but not with what we are?

Thirdly, plugging into an experience machine limits us to a man-made reality, to a world no deeper or more important than that which people can construct. There is no *actual* contact with any deeper reality, though the experience of it can be simulated. Many persons desire to leave themselves open to such contact and to a plumbing of deeper significance.¹ This clarifies the

Robert Nozick, "The Experience Machine," from *Anarchy, State and Utopia* (Basic Books, 1974), 42–5. Reprinted with permission of Perseus Books Group and Wiley-Blackwell.

intensity of the conflict over psychoactive drugs, which some view as mere local *experience machines*, and others view as avenues to a deeper reality; what some view as equivalent to surrender to the experience machine, others view as following one of the reasons *not* to surrender!

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. We can continue to imagine a sequence of machines each designed to fill lacks suggested for the earlier machines. For example, since the experience machine doesn't meet our desire to *be* a certain way, imagine a transformation machine which transforms us into whatever sort of person we'd like to be (compatible with our staying us). Surely one would not use the transformation machine to become as one would wish, and thereupon plug into the experience machine!² So something matters in addition to one's experiences *and* what one is like. Nor is the reason merely that one's experiences are unconnected with what one is like. For the experience

machine might be limited to provide only experiences possible to the sort of person plugged in. Is it that we want to make a difference in the world? Consider then the result machine, which produces in the world any result you would produce and injects your vector input into any joint activity. We shall not pursue here the fascinating details of these or other machines. What is most disturbing about them is their living of our lives for us. Is it misguided to search for *particular* additional functions beyond the competence of machines to do for us? Perhaps what we desire is to live (an active verb) ourselves, in contact with reality. (And this, machines cannot do *for* us.) Without elaborating on the implications of this, which I believe connect surprisingly with issues about free will and causal accounts of knowledge, we need merely note the intricacy of the question of what matters *for people* other than their experiences. Until one finds a satisfactory answer, and determines that this answer does not *also* apply to animals, one cannot reasonably claim that only the felt experiences of animals limit what we may do to them.

Notes

1. Traditional religious views differ on the *point* of contact with a transcendent reality. Some say that contact yields eternal bliss or Nirvana, but they have not distinguished this sufficiently from merely a *very* long run on the experience machine. Others think it is intrinsically desirable to do the will of a higher being which created us all, though presumably no one would think this if we discovered we had been created as an object of amusement by some superpowerful child from another galaxy or dimension. Still others imagine an eventual merging with a higher reality, leaving unclear its desirability, or where that merging leaves us.
2. Some wouldn't use the transformation machine at all; it seems like *cheating*. But the one-time use of the trans-

formation machine would not remove all challenges; there would still be obstacles for the new us to overcome, a new plateau from which to strive even higher. And is this plateau any the less earned or deserved than that provided by genetic endowment and early childhood environment? But if the transformation machine could be used indefinitely often, so that we could accomplish anything by pushing a button to transform ourselves into someone who could do it easily, there would remain no limits we *need* to strain against or try to transcend. Would there be anything left *to do*? Do some theological views place God outside of time because an omniscient omnipotent being couldn't fill up his days?

What Makes Someone's Life Go Best

Derek Parfit

What would be best for someone, or would be most in this person's interests, or would make this person's life go, for him, as well as possible? Answers to this question I call *theories about self-interest*. There are three kinds of theory. On *Hedonistic Theories*, what would be best for someone is what would make his life happiest. On *Desire-Fulfilment Theories*, what would be best for someone is what, throughout his life, would best fulfil his desires. On *Objective List Theories*, certain things are good or bad for us, whether or not we want to have the good things, or to avoid the bad things.

Narrow Hedonists assume, falsely, that pleasure and pain are two distinctive kinds of experience. Compare the pleasures of satisfying an intense thirst or lust, listening to music, solving an intellectual problem, reading a tragedy, and knowing that one's child is happy. These various experiences do not contain any distinctive common quality.

What pains and pleasures have in common are their relations to our desires. On the use of 'pain' which has rational and moral significance, all pains are when experienced unwanted, and a pain is worse or greater the more it is unwanted. Similarly, all pleasures are when experienced wanted, and they are better or greater the more they are wanted. These are the

claims of *Preference-Hedonism*. On this view, one of two experiences is more pleasant if it is preferred.

This theory need not follow the ordinary uses of the words 'pain' and 'pleasure'. Suppose that I could go to a party to enjoy the various pleasures of eating, drinking, laughing, dancing, and talking to my friends. I could instead stay at home and read *King Lear*. Knowing what both alternatives would be like, I prefer to read *King Lear*. It extends the ordinary use to say that this would give me more pleasure. But on Preference-Hedonism, if we add some further assumptions given below, reading *King Lear* would give me a better evening. Griffin cites a more extreme case. Near the end of his life Freud refused pain-killing drugs, preferring to think in torment than to be confusedly euphoric. Of these two mental states, euphoria is more pleasant. But on Preference-Hedonism thinking in torment was, for Freud, a better mental state. It is clearer here not to stretch the meaning of the word 'pleasant'. A Preference-Hedonist should merely claim that, since Freud preferred to think clearly though in torment, his life went better if it went as he preferred.¹

Consider next *Desire-Fulfilment Theories*. The simplest is the *Unrestricted Theory*. This claims that what is best for someone is what would best fulfil *all* of his desires, throughout his life. Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. Much later, when I have forgotten

Derek Parfit, "What Makes Someone's Life Go Best," from *Reasons and Persons* (Oxford University Press, 1984), 493–502. Reprinted with permission of Oxford University Press.

our meeting, the stranger is cured. On the Unrestricted Desire-Fulfilment Theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory.

Another theory appeals only to someone's desires about his own life. I call this the *Success Theory*. This theory differs from Preference-Hedonism in only one way. The Success Theory appeals to all of our preferences about our own lives. A Preference-Hedonist appeals only to preferences about those present features of our lives that are introspectively discernible. Suppose that I strongly want not to be deceived by other people. On Preference-Hedonism it would be better for me if I believe that I am not being deceived. It would be irrelevant if my belief is false, since this makes no difference to my state of mind. On the Success Theory, it would be worse for me if my belief is false. I have a strong desire about my own life – that I should not be deceived in this way. It is bad for me if this desire is not fulfilled, even if I falsely believe that it is.

When this theory appeals only to desires that are about our own lives, it may be unclear what this excludes. Suppose that I want my life to be such that all of my desires, whatever their objects, are fulfilled. This may seem to make the Success Theory, when applied to me, coincide with the Unrestricted Desire-Fulfilment Theory. But a Success Theorist should claim that this desire is not really about my own life. This is like the distinction between a real change in some object, and a so-called *Cambridge-change*. An object undergoes a Cambridge-change if there is any change in the true statements that can be made about this object. Suppose that I cut my cheek while shaving. This causes a real change in me. It also causes a change in Confucius. It becomes true, of Confucius, that he lived on a planet in which later one more cheek was cut. This is merely a Cambridge-change.

Suppose that I am an exile, and cannot communicate with my children. I want their lives to go well. I might claim that I want to live the life of someone whose children's lives go well. A Success Theorist should again claim that this is not really a desire about my own life. If unknown to me one of my children is killed by an avalanche, this is not bad for me, and does not make my life go worse.

A Success Theorist *would* count some similar desires. Suppose that I try to give my children a good start in life. I try to give them the right education, good habits, and psychological strength. Once again, I am now an

exile, and will never be able to learn what happens to my children. Suppose that, unknown to me, my children's lives go badly. One finds that the education that I gave him makes him unemployable, another has a mental breakdown, another becomes a petty thief. If my children's lives fail in these ways, and these failures are in part the result of mistakes I made as their parent, these failures in my children's lives would be judged to be bad for me on the Success Theory. One of my strongest desires was to be a successful parent. What is now happening to my children, though it is unknown to me, shows that this desire is not fulfilled. My life failed in one of the ways in which I most wanted it to succeed. Though I do not know this fact, it is bad for me, and makes it true that I have had a worse life. This is like the case where I strongly want not to be deceived. Even if I never know, it is bad for me both if I am deceived and if I turn out to be an unsuccessful parent. These are not introspectively discernible differences in my conscious life. On Preference-Hedonism, these events are not bad for me. On the Success Theory, they are.

Because they are thought by some to need special treatment, I mention next the desires that people have about what happens after they are dead. For a Preference-Hedonist, once I am dead, nothing bad can happen to me. A Success Theorist should deny this. Return to the case where all my children have wretched lives, because of the mistakes I made as their parent. Suppose that my children's lives all go badly only after I am dead. My life turns out to have been a failure, in one of the ways I cared about most. A Success Theorist should claim that, here too, this makes it true that I had a worse life.

Some Success Theorists would reject this claim. Their theory ignores the desires of the dead. I believe this theory to be indefensible. Suppose that I was asked, 'Do you want it to be true that you were a successful parent even after you are dead?' I would answer 'Yes'. It is irrelevant to my desire whether it is fulfilled before or after I am dead. These Success Theorists count it as bad for me if my desire is not fulfilled, even if, because I am an exile, I never know this. How then can it matter whether, when my desire is not fulfilled, I am dead? All that my death does is to *ensure* that I will never know this. If we think it irrelevant that I never know about the non-fulfilment of my desire, we cannot defensibly claim that my death makes a difference.

I turn now to questions and objections which arise for both Preference-Hedonism and the Success Theory.

Should we appeal only to the desires and preferences that someone actually has? Return to my choice between going to a party or staying at home to read *King Lear*. Suppose that, knowing what both alternatives would be like, I choose to stay at home. And suppose that I never later regret this choice. On one theory, this shows that staying at home to read *King Lear* gave me a better evening. This is a mistake. It might be true that, if I had chosen to go to the party, I would never have regretted that choice. According to this theory, this would have shown that going to the party gave me a better evening. This theory thus implies that each alternative would have been better than the other. Since this theory implies such contradictions, it must be revised. The obvious revision is to appeal not only to my actual preferences, in the alternative I choose, but also to the preferences that I would have had if I had chosen otherwise.²

In this example, whichever alternative I choose, I would never regret this choice. If this is true, can we still claim that one of the alternatives would give me a better evening? On some theories, when in two alternatives I would have such contrary preferences, neither alternative is better or worse for me. This is not plausible when one of my contrary preferences would have been much stronger. Suppose that, if I choose to go to the party, I shall be only mildly glad that I made this choice, but that, if I choose to stay and read *King Lear*, I shall be extremely glad. If this is true, reading *King Lear* gives me a better evening.

Whether we appeal to Preference-Hedonism or the Success Theory, we should not appeal only to the desires or preferences that I actually have. We should also appeal to the desires and preferences that I would have had, in the various alternatives that were, at different times, open to me. One of these alternatives would be best for me if it is the one in which I would have the strongest desires and preferences fulfilled. This allows us to claim that some alternative life would have been better for me, even if throughout my actual life I am glad that I chose this life rather than this alternative. [...]

Turn now to the third kind of Theory that I mentioned: the Objective List Theory. According to this theory, certain things are good or bad for people, whether or not these people would want to have the good things, or to avoid the bad things. The good things might include moral goodness, rational activity, the development of one's abilities, having children and

being a good parent, knowledge, and the awareness of true beauty. The bad things might include being betrayed, manipulated, slandered, deceived, being deprived of liberty or dignity, and enjoying either sadistic pleasure, or aesthetic pleasure in what is in fact ugly.³

An Objective List Theorist might claim that his theory coincides with the Global version of the Success Theory. On this theory, what would make my life go best depends on what I would prefer, now and in the various alternatives, if I knew all of the relevant facts about these alternatives. An Objective List Theorist might say that the most relevant facts are what his theory claims – what would in fact be good or bad for me. And he might claim that anyone who knew these facts would want what is truly good for him, and want to avoid what would be bad for him.

If this was true, though the Objective List Theory would coincide with the Success Theory, the two theories would remain distinct. A Success Theorist would reject this description of the coincidence. On his theory, nothing is good or bad for people, whatever their preferences are. Something is bad for someone only if, knowing the facts, he wants to avoid it. And the relevant facts do not include the alleged facts cited by the Objective List Theorist. On the Success Theory it is, for instance, bad for someone to be deceived if and because this is not what he wants. The Objective List Theorist makes the reverse claim. People want not to be deceived because this is bad for them.

As these remarks imply, there is one important difference between on the one hand Preference-Hedonism and the Success Theory, and on the other hand the Objective List Theory. The first two kinds of theory give an account of self-interest which is entirely factual, or which does not appeal to facts about value. The account appeals to what a person does and would prefer, given full knowledge of the purely non-evaluative facts about the alternatives. In contrast, the Objective List Theory appeals directly to facts about value.

In choosing between these theories, we must decide how much weight to give to imagined cases in which someone's fully informed preferences would be bizarre. If we can appeal to these cases, they cast doubt on both Preference-Hedonism and the Success Theory. Consider the man that Rawls imagined who wants to spend his life counting the numbers of blades of grass in different lawns. Suppose that this man knows that he

could achieve great progress if instead he worked in some especially useful part of Applied Mathematics. Though he could achieve such significant results, he prefers to go on counting blades of grass. On the Success Theory, if we allow this theory to cover all imaginable cases, it could be better for this person if he counts his blades of grass rather than achieves great and beneficial results in Mathematics.

The counter-example might be more offensive. Suppose that what someone would most prefer, knowing the alternatives, is a life in which, without being detected, he causes as much pain as he can to other people. On the Success Theory, such a life would be what is best for this person.

We may be unable to accept these conclusions. Ought we therefore to abandon this theory? [...] Suppose we agree that, in some imagined cases, what someone would most want both now and later, fully knowing about the alternatives, would *not* be what would be best for him. If we accept this conclusion, it may seem that we must reject both Preference-Hedonism and the Success Theory. [...]

It might be claimed instead that we can dismiss the appeal to such imagined cases. It might be claimed that what people would in fact prefer, if they knew the relevant facts, would always be something that we could accept as what is really good for them. Is this a good reply? If we agree that in the imagined cases what someone would prefer might be something that is bad for him, in these cases we have abandoned our theory. If this is so, can we defend our theory by saying that, in the actual cases, it would not go astray? I believe that this is not an adequate defence. But I shall not pursue this question here.

This objection may apply with less force to Preference-Hedonism. On this theory, what can be good or bad for someone can only be discernible features of his conscious life. These are the features that, at the time, he either wants or does not want. I asked above whether it is bad for people to be deceived because they prefer not to be, or whether they prefer not to be deceived because this is bad for them. Consider the comparable question with respect to pain. Some have claimed that pain is intrinsically bad, and that this is why we dislike it. As I have suggested, I doubt this claim. After taking certain kinds of drug, people claim that the quality of their sensations has not altered, but they no longer dislike these sensations. We would regard such drugs as effective analgesics. This

suggests that the badness of a pain consists in its being disliked, and that it is not disliked because it is bad. The disagreement between these views would need much more discussion. But, if the second view is better, it is more plausible to claim that whatever someone wants or does not want to experience – however bizarre we find his desires – should be counted as being for this person truly pleasant or painful, and as being for that reason good or bad for him. There may still be cases where it is plausible to claim that it would be bad for someone if he enjoys certain kinds of pleasure. This might be claimed, for instance, about sadistic pleasure. But there may be few such cases.

If instead we appeal to the Success Theory, we are not concerned only with the experienced quality of our conscious life. We are concerned with such things as whether we are achieving what we are trying to achieve, whether we are being deceived, and the like. When considering this theory, we can more often plausibly claim that, even if someone knew the facts, his preferences might go astray, and fail to correspond to what would be good or bad for him.

Which of these different theories should we accept? I shall not attempt an answer here. But I shall end by mentioning another theory, which might be claimed to combine what is most plausible in these conflicting theories. It is a striking fact that those who have addressed this question have disagreed so fundamentally. Many philosophers have been convinced Hedonists; many others have been as much convinced that Hedonism is a gross mistake.

Some Hedonists have reached their view as follows. They consider an opposing view, such as that which claims that what is good for someone is to have knowledge, to engage in rational activity, and to be aware of true beauty. These Hedonists ask, 'Would these states of mind be good, if they brought no enjoyment, and if the person in these states of mind had not the slightest desire that they continue?' Since they answer No, they conclude that the value of these states of mind must lie in their being liked, and in their arousing a desire that they continue.

This reasoning assumes that the value of a whole is just the sum of the value of its parts. If we remove the part to which the Hedonist appeals, what is left seems to have no value, hence Hedonism is the truth.

Suppose instead that we claim that the value of a whole may not be a mere sum of the value of its parts. We might then claim that what is best for people is a

composite. It is not just their being in the conscious states that they want to be in. Nor is it just their having knowledge, engaging in rational activity, being aware of true beauty, and the like. What is good for someone is neither just what Hedonists claim, nor just what is claimed by Objective List Theorists. We might believe that if we had *either* of these, *without the other*, what we had would have little or no value. We might claim, for example, that what is good or bad for someone is to have knowledge, to be engaged in rational activity, to

experience mutual love, and to be aware of beauty, while strongly wanting just these things. On this view, each side in this disagreement saw only half of the truth. Each put forward as sufficient something that was only necessary. Pleasure with many other kinds of object has no value. And, if they are entirely devoid of pleasure, there is no value in knowledge, rational activity, love, or the awareness of beauty. What is of value, or is good for someone, is to have both; to be engaged in these activities, and to be strongly wanting to be so engaged.

Notes

1. J. P. Griffin, 'Are There Incommensurable Values?', *Philosophy and Public Affairs*, 7/1 (Autumn, 1977).
2. See P. Bricker, 'Prudence', *The Journal of Philosophy*, 83/7 (July, 1980).
3. H. Sidgwick, *The Methods of Ethics*, (Macmillan: London, 1907), 111–12.