Published 2016 In: *The Routledge Handbook of Philosophy of Well-Being*, (ed.) Guy Fletcher (New York: Routledge).

The Experience Machine and the Experience Requirement

Jennifer Hawkins
Duke University

One particular thought experiment—Robert Nozick's experience machine—has had a huge impact on the way philosophers think about well-being.[1] Indeed, many assume it completely refutes hedonism once and for all, and not merely hedonism, but any theory that focuses exclusively on mental states. However, as we shall see, Nozick's example and its implications are more complex than people typically realize. The original example goes like this:

> 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 experiences?[2]

In essence, Nozick asks us to imagine the possibility of a machine capable of giving someone *any* experience she might want. In more contemporary terms, we could think of it as the most powerful virtual reality machine ever conceived. The machine stimulates all of the brain's sensory input channels, providing experiences as phenomenologically rich as any in real life. For example, it could give someone the experience of skiing down a snowy mountain complete with vision of mountains, snow and trees, the feel of wind on her face, and the bodily sensations of gliding

---

[1] I treat 'well-being' and 'welfare' as synonyms. I assume that theories of well-being (or of welfare) are about a special kind of value, the kind under discussion when we discuss what is good *for* a particular person. I also sometimes refer to this kind of value as 'prudential value' and occasionally use the adjective 'prudential' to signal a focus on reasons relevant to a particular person's good.
[2] Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), 42-5 and *The Examined Life* (New York: Simon and Schuster, 1989), 104-8.

smoothly and swiftly downward. Indeed, we are to imagine that the machine is so good that, from within, it is *impossible* to tell the difference between real experiences and machine produced ones. It is also important to note that once someone enters the machine, the machine ensures that she forgets where she is and how her experiences are being crafted. She believes her experience is real, even though it is caused by electrodes attached to her brain.

Nozick expresses confidence that most people would not want to plug in. However, *if* the quality of experience is all that matters in a life, then it seems that one ought to want to plug in, since the machine is, by hypothesis, the best way to ensure large quantities of high quality experience. Interestingly, this is true no matter how you define 'good' experience. I shall use the label 'experientialism' for any theory that defines well-being purely in terms of mental states, i.e. any theory that says only experiential states can be bearers of intrinsic welfare value. Hedonism is simply one form—albeit the most familiar—of experientialism. Although Nozick's original target was hedonism, the thought experiment, *if* it works, works equally well against any form of experientialism. Many philosophers take the example to show both that ordinary people do not think about welfare in (exclusively) experientialist terms, and that the correct theory of well-being—whatever else it is—is not experientialist.

Despite the apparent simplicity of this thought experiment, the issues it raises are complex and relatively under-explored. The aim of this chapter is to rectify that. I begin by considering how the experience machine differs from other common objections to hedonism. I take a closer look at the structure of the argument it is supposed to provide against experientialism. In particular, I highlight some of the confusions and problems that arise from the specific way Nozick sets up his thought experiment. I then consider whether it is possible to re-formulate the example in a way that avoids these problems. I next consider the question: what would follow if we *did* reject experientialism? As we shall see, there would still be much to decide about which non-experientialist theory of well-being to accept. Finally, I consider the relationship between rejecting experientialism (as Nozick

hopes we will do) and rejecting what has come to be known as "the experience requirement," explaining why these are not precisely the same thing.[3]

§1.0 A Distinctive Kind of Objection

The original target of Nozick's thought experiment is hedonism, a view about well-being according to which the only thing intrinsically valuable (from the prudential point of view) is pleasure and the only thing intrinsically bad (from the prudential point of view) is pain. Hedonism aims to tell us something quite general about what makes lives better or worse.[4]

One prominent, traditional strategy of critics of hedonism is to find fault with hedonism's account of valuable mental states. The basic aim of such an objector is to establish that there are more types of valuable consciousness than simply pleasure (and more types of bad consciousness than simply pain). How successful any such objection is, depends partly on one's views about what is valuable in conscious experience and partly on how elastic one is willing to be in one's definition of terms such as 'pleasure' and 'pain.' A few examples may make this clearer. John Stuart Mill famously defined 'happiness' in terms of pleasure and the absence of pain.[5] But various people, over time, have objected to his simple equation of happiness with pleasure. Even assuming that 'happiness' is the name for a psychological state, many have claimed it is the name for a distinct psychological state—one that is both more complex and more valuable than mere pleasure.[6] If one were to adopt such a view of happiness and combine it with the claim that well-being consists of happiness, one would be defending a version of experientialism. It would not, however, deserve the

[3] The phrase "experience requirement' originates with James Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Clarendon Press, 1986), p13.
[4] See also Alex Gregory, "Hedonism," ch.9 of this volume.
[5] John Stuart Mill, [1861], *Utilitarianism* (New York: Barnes and Noble, 2005), p. 7.
[6] Two prominent examples of theorists who reject the equation of happiness with pleasure in favor of more psychologically complex accounts of happiness are L. W. Sumner, *Welfare, Happiness, and Ethics* (Oxford: Oxford University Press, 1996) and Daniel M. Haybron, *The Pursuit of Unhappiness* (New York: Oxford University Press, 2008). Though both authors are deeply interested in the nature of happiness, their respective accounts are quite different. Importantly, neither is an experientialist, since neither accepts a simple equation of happiness with well-being.

label 'hedonism' because of the explicit rejection of the idea that pleasure is the major welfare value.

Some objectors in this category go even further and argue that among the valuable types of consciousness are some painful or unpleasant states. For example, if we sometimes care more about the *process* of thinking or about the *contents* of our thoughts than about how we feel, we might in some instances reasonably prefer sensory pain over sensory pleasure despite the fact that traditional hedonism would view such a preference as prudentially irrational. James Griffin offers the example of Sigmund Freud, who during his final illness preferred to think in torment without pain medications given that the medications dulled his thoughts.[7] If we think Freud's choice makes prudential sense, then this suggests we do not accept the traditional hedonist characterization of valuable consciousness. However, in itself, it does not challenge the basic idea that internal mental experience is what matters. After all, according to the story, Freud tolerated pain for the sake of *thinking*.

Nozick's thought experiment has gained so much attention precisely because it departs radically from this familiar type of criticism, and instead offers a critique of experientialism in all its forms. Whereas traditional objectors focused on the idea that there are more types of valuable consciousness than just pleasure, Nozick's example is meant to establish that there is more to well-being than valuable consciousness *however* one chooses to define "valuable consciousness."

It is worth noting, that although the experience machine is the example used most often to attack experientialism, there are a number of other, closely related examples in the literature on well-being that are intended to make a similar point. These typically don't involve a machine, but simply posit deception or ignorance such as might arise in the ordinary course of living. And the person in the example is not lacking all or even most knowledge of her life, but simply knowledge of one or more key aspects. For example, Sumner describes a case in which someone is happily involved in a relationship, but doesn't know that her partner is unfaithful.[8]

[7] Griffin, *Well-Being*, p8.
[8] Sumner, *Welfare, Happiness, and Ethics*, 157.

Scanlon uses the example of someone who is secretly despised by those he falsely thinks of as friends.[9] And still other theorists appeal to examples in which someone happily believes she has accomplished something when she hasn't really.[10] The differences are less important than the similarities, however. For as with Nozick's example, the point is to elicit the intuition that something in these lives is not good, or at least not as good as it could or should be, and this despite the fact that the agents in question are happy in their delusions: a conclusion a hedonist cannot accept.

§2.0 Problems with the Argument

Despite its fame, the experience machine example can be very confusing. Because it is a thought experiment, we are supposed to draw conclusions on the basis of our own intuitive reactions to the case. Nozick is clear that he thinks most people will *not* want to sign up for life in the machine. But is he right? And what really follows if he is?

Insofar as there is an argument, it seems to be *roughly* this:

(1) If some form of experientialism is true, most people will, upon encountering the thought-experiment, want to sign up for the machine.
(2) In fact, most people who encounter the example want *not* to sign up for the machine.
(C) Therefore, no form of experientialism is true.

Let me begin with some remarks about premise (2). Nozick writes as if he is confident that no one (or almost no one) would want to sign up for a life in the

[9] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 112.
[10] Russ Shafer-Landau, *The Fundamentals of Ethics*, 2nd edn (New York: Oxford University Press, 2012), 53. Shelly Kagan gives a single example involving all three elements, namely a man deceived about spouse, friends, and accomplishments. Shelly Kagan, *Normative Ethics* (Boulder, CO: Westview Press, 1998), 36.

machine. However, we don't really know whether that is correct. Philosophers sometimes write and talk as if it is a well-known fact that most people do not want to sign up. But that is an empirical question, and one that (to my knowledge) has never been rigorously tested. Of course, there is lots of anecdotal evidence from philosophers who have taught the example over the years. But the anecdotal evidence is mixed, and all sorts of factors may contribute to the replies students give. Classrooms are hardly controlled environments. So we just don't know how most people would respond.[11]

Nonetheless, it is natural to wonder: *If* he were right, and most people did not want to sign up, would that demonstrate that experientialism is false? Not necessarily. In fairness, there is an important core truth in the way the example is set up. But other features of Nozick's presentation make it difficult to draw any clear conclusions.

The core truth, which is worth stating, is just this: *if* hedonism or some other version of experientialism were true, then assuming the machine really is as powerful as claimed, it would make most sense (from a purely prudential point of view) to sign up. This is because the machine would be able to give a person *the best life possible*. No other option would be as good. Some people claim that real life—at least in theory—could compete with the machine. For example, if we assume that pleasure is what matters, then the claim would be that it is at least possible for a real life to contain as much pleasure as a machine life. If that were the case, then an extremely pleasurable life might be tied with machine life for best. But although this isn't logically ruled out, it is extremely unlikely. Moreover, since even in that scenario no life is *better* than the machine life, and since machine life is so much more dependable than real life, the machine would clearly be the better prudential choice for any given individual.

[11] Though I am not aware of any rigorous, empirical testing of Nozick's original example, De Brigard has tested people's reactions to scenarios in which they learn they are already hooked up to an experience machine and have to decide whether or not to leave it. The results are interesting, but for reasons explained later on they don't tell us anything about the truth or falsity of *experientialism*. Felipe De Brigard, "If You Like It, Does It Matter If It's Real?" *Philosophical Psychology* 23: 1: (2010): 43-57.

However, the argument requires people to recognize this fact and *then* make a decision about whether to sign up *based purely on considerations about their own welfare*. Now given that not every motive a person has for doing something is a motive related to her own welfare, this immediately raises the question of how to distinguish reasons of self-interest from other types of reasons. This is important because it is plausible to think that various welfare irrelevant reasons may influence the choice people make, either consciously or unconsciously. But if other motives are at work then premise (1) which states that, "If some form of experientialism is true, most people will want to sign up for the machine" might be false. Experientialism might be true even though most people do *not* wish to sign up. Unless we can confidently rule out the influence of such reasons, which requires that we first be able to reliably identify them, we can't interpret lack of willingness to sign up as indicative of the truth or falsity of experientialism.

In the literature one can find many different expressions of the same basic concern, namely that people may refuse to sign up for reasons other than having rejected the thesis that it is prudentially good to do so. Many people have found it difficult to really grasp and take seriously a possibility so remote from real life. Even though technology is more sophisticated now than when Nozick wrote the example, it is still a long, long way from being able to substitute plausibly for all of our five senses, much less for any length of time. Thus, it can be hard to give credence to the idea that a machine might really be that powerful, and this might make us reluctant to sign up. In a similar vein, it can be hard to put aside worries that the machine might malfunction, or might fail to deliver the best possible experiences. As part of the thought experiment we are supposed to assume it won't malfunction, but how could we ever know that about any real machine?[12] As we shall see in the next section there are also credible worries about unconscious motives such as status quo bias.[13]

[12] Worries like these are expressed by Sumner, *Welfare*, *Happiness, and Ethics,* p.95.

[13] Jason Kawall, "The Experience Machine and Mental State Theories of Well-Being," *Journal of Value Inquiry* 33: 3: (1999): 381-87; De Brigard, "If You Like It."

Many of the problems arise from the fact that Nozick presents the example as a *choice* for the reader. We are asked whether we—who are, by hypothesis, not now living in a machine—would agree to sign up for life. This puts us in a very funny position. It is stipulated that in the machine we will have great experiences of whatever type we value. Moreover, we will not—once in the machine—know that our experiences aren't real. But of course, as we contemplate whether to sign up, we know that future experiences in the machine will not be real. And because this invites all sorts of welfare irrelevant reasons to come into play, it creates problems.

People can desire things other than their own welfare, and sometimes these desires are strong enough to lead them to act in ways that are not welfare maximizing. Experientialism in itself doesn't rule this out. It is just a theory about what is good for us, and it could be a true theory about our good even if we do not always choose what is good for us.[14] For example, people can have purely altruistic desires, desires for the good of another person. If that is possible, then a person might not want to sign up because by doing so she would make it the case that she could no longer help others. After all, once in the machine she would no longer really be interacting with other people, just computer simulations of people. Anticipating this particular kind of worry, Nozick stipulated that part of the thought experiment should include imagining that others are well off and not in need of our help.[15] But while that might handle purely altruistic desires, these are not the only potentially problematic desires.

Consider the fact that many people have a strong, brute desire to know things, a desire that is not obviously welfare related. Though we talk about curiosity killing the cat, we invented that expression to talk about *ourselves*. It points to the idea that there is a stubborn quality to this particular human desire, that people often desire to know things even when it is not good for them to know. Precisely because entering the machine requires us to give up all knowledge, it is plausible to

[14] Kawall, "Experience Machine"; Matthew Silverstein, "In Defense of Happiness: A Response to the Experience Machine," *Social Theory and Practice* 26:2: (2000): 279-300; Sharon Hewitt, "What Do Our Intuitions About the Experience Machine Really Tell Us About Hedonism?" *Philosophical Studies* 151:3: (2010): 331-49.
[15] Nozick, *Anarchy*, 43; *Examined Life*, 105.

think that people might balk at the idea regardless of whether it would be good for them to enter. In my own case, at least, I know I would be unwilling to enter the machine, because it would entail not knowing what happens to those I love. Indeed, I would go as far as to claim that part of what it is to love someone is to want to know what happens to them. Of course, the primary desire of one who loves is the desire for the welfare of the loved one. But one also wants to see the other's life unfold, to track the loved one's progress through the world. It would be small comfort simply to be assured that my loved ones will be okay if I enter the machine. I would still understand that a choice to enter is a choice to forgo any further knowledge of these people. The issue, of course, is about what such reluctance *means*. I admit that my own sympathies are not experientialist, so I tend to assume that (in most cases at least) knowledge of the sort that matters to me is also good for me. But in fairness to experientialists, I am also pretty sure that my desire to know has no grounding in, and is not limited by, facts about my welfare: that I would still want to know *whether or not it was good for me*. It seems plausible that many people have similarly strong, welfare-independent strands of curiosity. Suppose now that it turns out that many people do not want to sign up for the experience machine, and they cite as their reason a desire to know how things really are in the world. Unless we can rule out the possibility that these desires are welfare irrelevant desires, we cannot draw any conclusions about experientialism from the fact of their reluctance.

In short, the example as formulated is unable to escape from a certain kind of dilemma. On the one hand, if we had some reliable way of stipulating ahead of time which desires are self-interested, we might be able to show that people were rejecting the machine for self-interested reasons, which is what the argument against experientialism needs. However, we can only have such a distinction if we already have a theory of well-being. It simply begs the question against the experientialist to begin with such a stipulation. On the other hand, without it, it will in many cases be unclear what to conclude even if, as predicted, many people don't want to sign up.

§3.0 Just How Bad is Machine Life?

Another problem with Nozick's example is that it invites a certain kind of misreading, or (if not literally a misreading) at least a conflation of issues. For many people assume that the point of the example is to persuade us that we should *never for any reason* sign up for the machine. Certainly some of what Nozick says in his original presentation suggests that interpretation. But it is not necessary to accept this strong claim in order to reject experientialism. A non-experientialist can consistently grant that it *sometimes* makes sense to sign up for the machine. The example thus conflates the project of rejecting experientialism and the project of defending a strong view about the intrinsic value of connection with reality.

To see the problem more clearly, it can help to think of theories of well-being as giving us rankings of possible lives. Obviously a theory of *well*-being aims to tell us what makes *good* lives *good*. But ideally it should also tell us what makes bad lives bad, and which possible lives are in the middle and why. It should give us insight into those features of lives that make them better or worse, and so enable us—at least in theory—to rank possible lives from best to worst.

Hedonists rank lives according to a total score, reached by adding up pleasure, adding up pains, and subtracting the pain from the pleasure. A positive net score (more pleasure than pain) is good, but the best life is a life of maximal pleasure and no pain, and the worst would be a life of maximal pain and no pleasure. Different experientialist theories will, of course, produce different rankings, but the approach to ranking will be similar. As we saw in the last section, the important truth about the experience machine is that *if* some version of experientialism is true and *if* we grant that the machine really is as powerful as it is claimed to be, then life in the machine represents the best possible life, or at least the best possible life choice.

To reject experientialism is to reject the idea that machine life is *best*. But notice that this is still a far cry from claiming that machine life is bad or even worst.

Among those who reject the idea that machine life is the best life, there could still be lots of disagreement about where precisely in the ranking of possible lives machine life falls. Only the extreme claim that machine life is the *worst* possible life would support the claim that it never, no matter the alternatives, makes sense to sign up. Indeed, many theorists who are not hedonists allow that happiness is a significant, intrinsic prudential good.[16] But if that is true, then machine life will be better than some possible alternative lives entirely lacking in happiness.

In his second, later discussion of the experience machine in *The Examined Life*, Nozick is explicit that the proper question is whether machine life is *best*. He writes, "The question is not whether plugging in is preferable to extremely dire alternatives—lives of torture, for instance—but whether plugging in would constitute the very best life, or tie for being best" (105). However, even though he makes the point, he undermines its strength by offering only one possible example of a life worse than machine life: a life of torture! So it is not surprising that this point is often lost. Many discussions of the experience machine still assume that the point of the example is to establish that machine life is very bad.

This is important because it speaks to a frequent reaction people have to Nozick's example. As we have seen, people interpret him as holding that it is always better to be outside the machine. Many of my students initially respond by insisting that whether it makes sense to sign up must depend on the alternatives. Perhaps for a homeless orphan living in a slum in one of the poorer countries of the world—someone with little hope of improving her situation—the experience machine would be a good option. As far as it goes, their point is reasonable. Even if Nozick would disagree (and that is not clear), many other non-experientialist philosophers would agree.[17] The important point is just to see that this response to Nozick is not a defense of experientialism. Even if Nozick himself ranks machine life low, the

[16] Indeed, Nozick himself emphasizes the value of happiness and laments the way in which the claims of hedonism distort our understanding of the nature and value of happiness. *Examined Life*, 99.
[17] Though again, it is important to emphasize that it is hard to know whether Nozick would disagree or not. He clearly invites his readers to place intrinsic value on having a connection with reality (which itself is open to several interpretations). But how much value this has, and how much disvalue is attached to its lack, is never worked out in detail.

experience machine example undermines experientialism (if it does) by suggesting that machine life is not best.

In an interesting set of empirical studies De Brigard presented students with scenarios in which they were told that they have been living in an experience machine. They were to imagine that all of their memories were produced by the machine, though they presumably once had a life out of the machine that they don't remember but which they could return to. They are given the option of doing precisely that: the option of returning to real life. De Brigard used different versions of the scenario. In one version no information is given to suggest anything about what the real life would be like. In the other two versions information about real life is given (in one case suggesting it is not good, in the other case suggesting it is good). The results were quite divided, but were definitely sensitive to the information about how good or bad the "real" life was.[18]

De Brigard takes it as a starting point of his paper that most people presented with Nozick's case do *not* want to sign up. He then sees himself as looking for an explanation of the *dual* fact that when people contemplate signing up they are reluctant to do so, but when people are asked to contemplate getting out, they are also reluctant to do so. He offers an interesting hypothesis in terms of status quo bias, the idea, well-established in psychology, that people are exceedingly cautious about giving up what they have. People have a tendency to overvalue what they already possess or what they already know. Given this tendency, an alternative must be viewed as considerably better than the status quo in order to motivate people to make a change.

There are two points I wish to make about this. First, *even if* we could draw a straightforward conclusion from De Brigard's results, the conclusion, though interesting, would not tell us anything useful about *experientialism*. By straightforward conclusion, I mean the conclusion that would be suggested if we could be sure that nothing other than welfare relevant considerations were contributing to choice. De Brigard's examples are intended to test the view that

[18] De Brigard, "If You Like It."

machine life is one of the worst possible lives. If it were true that most people believed this, then one would expect people who are told that they are in an experience machine to want to come out. Since they did not all want to leave the machine, this suggests that people do not all see machine life as the worst possible life, or even as particularly bad. It all depends on the alternatives. However, even if De Brigard's results could be read as showing this (and I don't think even he thinks they can, because of probable status quo bias), it would not tell us about the truth or falsity of *experientialism*. This is because, although showing that machine life is not the worst life might be an interesting reply to Nozick, it doesn't speak to the issue of whether machine life is best.

One might counter that if machine life is best, no one should have wanted to leave. But in De Brigard's example, unlike Nozick's, machine life was not characterized to make it clearly best, for in De Brigard's example, machine life is simply the life the person has lived up until now, which, like most lives, has both good and bad elements.

Second, and more importantly, if his hypothesis about status quo bias is correct, then it is hard to know what to conclude. I refer interested readers to the details of De Brigard's article. But in general, I think that the combined lesson of the last two sections is that setting up machine examples in terms of personal choice allows too many irrelevant factors to enter in. I want now to consider whether it is possible to reformulate the example to isolate intuitions about *experientialism*.

§4.0 A Reformulation

Is there a way to reformulate the example, so that it does a better job of isolating the relevant intuitions: intuitions that would distinguish experientialists from non-experientialists? Whether or not it solves all the problems, the following—from Roger Crisp—strikes me as a significant improvement.[19]

[19] Although I have taken the liberty of presenting the example with my own details, the basic idea of comparing two whole lives that are qualitatively identical although one is lived in a machine and one

Consider two twin girls, Molly and Polly. Imagine that Molly is born and has a great life in the real world. Readers can fill in the details of the life in whatever way is likely to make it seem attractive. This way we ensure that her life is qualitatively good. And let us imagine that she lives to a ripe old age of 100, ensuring her life is quantitatively good as well. Polly, her identical twin, is born a few minutes later, but Polly is immediately whisked away by the same superduper neuropsychologists Nozick describes, who hook her up to an experience machine. Inside the machine Polly lives a life that is qualitatively *identical* moment for moment to Molly's life. Whatever Molly really does, Polly has a virtual experience that is—from the inside—indistinguishable. Like Molly, Polly also lives for 100 years and then dies content, never knowing that her life was unreal. What we then ask ourselves is this: do we think that their lives are equal in prudential value or do we think that one of them has a better life than the other? An experientialist should say the lives are equally good. But a non-experientialist will think that Molly's life is a better life, even if neither Molly nor Polly is positioned to make this assessment.

Framed this way, the example escapes many of the earlier concerns. For one thing, worries about how to imagine such a powerful machine have less traction, since we don't worry about the future. We are simply told what the life was like and that it has already occurred, which somehow seems easier to believe or grasp, precisely because it is more determinate. Similarly, worries about machine malfunction seem to evaporate from this perspective, since we are no longer peering into an uncertain future for ourselves, but contemplating a completed life where it is just stipulated that the machine did not malfunction. We are simply told (and we fairly easily accept) that the machine gave Polly a life qualitatively identical to the one lived by Molly.

Most importantly, since no one is asked whether she wants to sign up, there is no room for welfare independent desires (ours or Polly's) to distract us from the primary question. Polly never makes a choice and neither do we. And because of

is lived in the real world is from Roger Crisp, *Reasons and the Good* (Oxford: Oxford University Press, 2006), 117-19.

this, we can more easily focus on our intuitions about the goodness of her life. We do not have to face all the problems that come from thinking about what it would mean to give up the life we have already begun, the life we are already invested in. We may be prone in our own case to status quo bias. But if so, that will not be triggered here. Nor will our own welfare unrelated desires get in the way.

Instead, we just have to decide whether Polly's life is lacking something that Molly's has. Finally, because the reformulation stipulates that both lives are enviably good from the inside, no distracting issues about ranking arise. Even if one thinks that Polly's life is worse than Molly's, one might also think that Polly's life is better than the real life of someone who is desperately poor, ill, and alone. In short, this version doesn't invite the conclusion (as Nozick's discussion seems to) that machine life is *never* choiceworthy. It forces us to focus on the narrower question of whether a good real life is better than an experientially good machine life.

§5.0 States of Affairs vs. Knowledge

Suppose for the moment that we think Polly's life is worse than Molly's. What precisely does this show? There are (at least) two ways of explaining the difference in value, and the literature on these issues does not typically make this clear.[20]

First, someone might think that what matters in life are the facts about what really happens. More precisely, we might think it matters which states of affairs come about. If we take this approach, we need some way of identifying which states of affairs matter: which states of affairs are significant and relevant to the value of this person's life. Desire theory uses (some of) an individual's desires to pick out the relevant states. According to desire theory, if I desire to accomplish some goal G, then what has value for me is the coming to be of the state of affairs in which I actually accomplish G. Usually, of course, when such states of affairs come about, I know this. But on the first view under consideration, knowledge is not *required* in

[20] Jennifer Hawkins, "Well-Being: What Matters Beyond the Mental?" *Oxford Studies in Normative Ethics*, Vol. 4, January 2015.

order for a state of affairs to have positive (or negative) prudential value. A person's life could thus be better than she thinks or worse than she thinks. I shall call theories like this—that accord value directly to states of affairs—SA theories, for *prudential value of states of affairs*. It is important to remember that desire theory is only one, albeit the most famous, example of an SA theory.

A very different, alternative conclusion one might reach emphasizes the prudential value of knowledge or some other positive epistemic relation such as true belief or justified true belief. For simplicity, I'll just discuss knowledge. On this view, knowledge about the facts of my life has positive prudential value for me. Again, of course, a theorist drawn to this idea will need a way of saying which things it is good to know. Presumably not all knowledge has value: I, at any rate, see little prudential value in knowing the number of ants living in my backyard! Precisely because knowledge is a *relation* between mind and world, it is not purely mental. So it is the kind of thing that Molly might have and Polly lack, even though their lives are experientially identical. I shall call theories like this—that accord value to epistemic relations—ER theories, for the *prudential value of epistemic relations*.

SA and ER are very different, and offer competing explanations of why Polly's life is worse than Molly's. Inside the experience machine Polly lacks knowledge. Most of her beliefs are false, even though she doesn't know this. And so an ER theory would see less value in her life than in Molly's. But notice as well that most of the significant facts of her life are not as she wants them to be either. Using the desire theory as an example of a SA theory, let us suppose that Polly (like Molly) at one point wishes to visit Japan. Whereas Molly actually visits Japan, Polly merely has virtual experiences that are Japan-like. Though she doesn't realize it, her desire is frustrated, not satisfied. Indeed, presumably most of Polly's significant life desires are frustrated, making her life quite bad from the standpoint of a desire theory. If we think that Polly's life is worse than Molly's the interesting question is: *Why?* Is it because Polly is so ignorant of the truth about her life? Or is it because the facts are not as she wants them to be? Or is it both?

To illustrate vividly the difference between SA and ER, consider the following four possible lives. Suppose that these four different scenarios occur in lives that are otherwise identical in every way, so that any difference in the value of these lives must be traceable to differences in these cases.

Life 1: Polly has a desire to G, her desire is frustrated, and she knows this.
Life 2: Polly has a desire to G, her desire is satisfied, and she knows this.
Life 3: Polly has a desire to G, her desire is frustrated, though she never knows this.
Life 4: Polly has a desire to G, her desire is satisfied, though she never knows this.

A desire theorist will rank these lives as follows: Lives 2 and 4 are equal in value and both are better than either 1 or 3 (which are also equal in value). Someone who accepts an ER theory that accords no direct value to states of affairs, will instead say that lives 1 and 2 are equal in value and both are better than either lives 3 or 4 (which are also equal in value). Of course, many plausible non-experientialist theories of well-being may allow that *both* states of affairs and epistemic relations are important. One does not have to accept one and reject the other. The point of doing so here is just to illustrate, as dramatically as possible, that they really are different theses. It is no doubt also true that many plausible non-experientialist theories of well-being will accord intrinsic value to things *other than* states of affairs and epistemic relations. For example, many theories will accord happiness some, though not exclusive, weight. If that's correct, then rankings will be complicated in more ways than illustrated here.

Still, it is worth emphasizing the difference between SA and ER, if only because, historically, philosophers have tended to overlook ER and other alternatives to a pure SA theory. According to one very popular and familiar story about the development of theories of well-being, the obvious solution to the problem posed by the experience machine is to adopt a desire theory. But while it is true that desire theory, which is a pure SA theory, is an alternative to experientialism, it is not the only one. Nor is tacit acceptance of desire theory the

only explanation of the intuition that Polly's life is worse than Molly's. That intuition by itself only tells us to reject experientialism. But once you do, there are various alternative views to choose from.

§6.0 Experientialism and the Experience Requirement

James Griffin coined the phrase "experience requirement" in the course of talking about the move from experientialism to desire theory. Whereas experientialism embraces, desire theory rejects, "the experience requirement." But what precisely is the experience requirement?

Following Griffin, when people talk about the rejection of the experience requirement they typically have in mind a theory that goes beyond the mental in a very strong sense. They typically have in mind a theory that gives no necessary role to mental states—a theory like a desire theory that assigns intrinsic value only to states of affairs, and only indirectly and contingently to mental states if these happen to be constituents of desired states of affairs. For example, a person can desire the state of affairs in which she is happy or the state of affairs in which she knows things. When that occurs, mental states figure indirectly in the account of welfare. But on such a view, there is no necessity that prudential goods or bads be experienced by the person who is thus made better or worse off.

Having said this, it is important to note that there is some disagreement in the literature about what it means to reject or, alternatively, incorporate, an experience requirement. Some philosophers assume that any theory that makes good experiences necessary for welfare is a theory that incorporates an experience requirement. Alternatively, and more in keeping with Griffin's usage, an experience requirement could be understood as the requirement that anything that affects welfare (positively or negatively) must enter experience. These two can come apart.

The theory of L. W. Sumner is a case in point.[21] According to Sumner, welfare is authentic happiness, where this phrase requires explanation. First, happiness is understood as a complex psychological state. It involves both judging one's life to be good *and* feeling good. As such, happiness for Sumner has both cognitive and affective dimensions. However, the theory is a hybrid theory in the sense that it also has non-mental requirements. Although happiness is necessary for welfare, it is not sufficient. In addition, Sumner imposes an authenticity condition, which in turn has two parts. I will not go into great detail about these, but they entail that a person who is happy can nonetheless be worse off than she thinks if either (a) her happiness depends upon false information, or if (b) her happiness is based on values that are not authentically hers.

The interesting feature of Sumner's view is its asymmetry: a person can be worse off than she thinks she is, but she cannot be better off than she thinks she is. Happiness is necessary for a good life. Since you know you are happy if you are happy, you are either doing as well as you think, or (if your happiness fails the external conditions) doing worse than you think. This is clearly a theory that gives a central role to experiential states. If we assume that to have an experience requirement simply means to make certain kinds of experience necessary for a good life, then Sumner's theory incorporates an experience requirement. And this appears to be Sumner's own understanding of the idea, since he describes himself as building the experience requirement back in.

However, if we consider Sumner's view in light of the second definition of experience requirement, we can see that it doesn't build in an experience requirement. For Sumner doesn't insist that in order to have an effect on welfare something must be experienced. Certain kinds of negative facts, which if known would undermine happiness, can without actually undermining happiness, make a person's life worse than she thinks it is. This is to allow that certain states of affairs outside awareness can nonetheless have an impact on welfare. In this sense Sumner's view does not incorporate an experience requirement.

[21] Sumner, *Welfare, Happiness, and Ethics*.

As is often true in philosophy the really important point is not which definition we adopt, but that we see the difference and track it in our theorizing. However, since I think more people tend to understand the experience requirement as the idea that something must be experienced if it is to have a positive or negative impact on welfare, I suggest to the profession that in future we adopt this definition. We must then simply keep in mind that it is possible for a theory to give great intrinsic weight to experience without incorporating an experience requirement.

What then of the relationship between experientialism and the experience requirement? To reject experientialism one must think that at least some of the bearers of intrinsic welfare value are non-mental. But it is possible to reject experientialism and still assign a big role in one's theory to experience (as Sumner does). And it is even possible to reject experientialism without rejecting the experience requirement at all. For it is possible to hold a view like the one I have elsewhere called the conditional value thesis, which maintains that the intrinsic bearers of welfare value are states of affairs, but insists that these have value for a person only if they are known.[22] Assessing whether or not such a view has any plausibility is beyond the scope of this chapter, and although I have described it elsewhere I do not defend it there. I mention it here simply to underscore the point that the rejection of experientialism and the rejection of an experience requirement are not the same thing.

[22] Hawkins, "What Matters Beyond the Mental?"