

In Defense of Happiness: A Response to the Experience Machine

Introduction

Hedonism has seen better days. As a philosophical doctrine, it flourished in the utilitarian climate maintained by the great nineteenth-century British empiricists Jeremy Bentham, John Stuart Mill, and Henry Sidgwick. And even though utilitarianism continues to hold sway among many contemporary ethical theorists, its hedonistic foundations have been all but abandoned. Nowadays, most philosophers entertain the idea of hedonism just long enough to dismiss it.

Traditionally, hedonism has come in many guises. The term itself is derived from the Greek word ἡδονή (*hēdonē*), meaning pleasure, and theories are hedonistic in virtue of the central role they assign to the notion of pleasure. Psychological hedonism is a view about human motivation. It holds that pleasure is the only possible object of a desire or pursuit. Moral hedonism contends that pleasure is the only thing one ought to desire or pursue. A third version, rational hedonism, maintains that pleasure is the only thing one has reason to desire and pursue.¹ The most familiar formulation of hedonism, though, and the one with which I am primarily concerned here, is a doctrine about well-being and prudential value.

Well-being is itself an elusive concept in need of explanation. An exhaustive and reductive definition is, however, beyond the scope of this essay; I offer instead some synonymous expressions. Although these expressions may not dispel all confusion, they will, I hope, help us recognize what is ascribed to a person when we say that he or she has well-being. Well-being is the general condition one has when one is faring well. It concerns how well a life is going at a particular moment and how well it is going "*for the individual whose life it is.*"² As I take it, then, well-being is synonymous

¹These definitions are taken from J.C.B. Gosling, "Hedonism," in Ted Honderich (ed.), *The Oxford Companion to Philosophy* (Oxford: Oxford University Press, 1995), p. 337.

²L.W. Sumner, *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press, 1996), p. 20. Properly speaking, then, one's level of well-being may fluctuate over even a short span of time. My life may be going well at one moment but then poorly at the next, or vice versa. We

with a person's own good, welfare, or interest. To say that something contributes to one's well-being is simply to say that it is in one's interest, that it is good for one, or that it makes one's life go better. Understood as such, well-being is an inherently evaluative concept: when we say that something promotes one's welfare, we ascribe a certain type of value to that thing. In particular, we identify it as having prudential value.³ Prudential value is conceptually distinct from other modes of value, such as aesthetic, perfectionist, or moral value. Unlike these other modes, prudential value is concerned with the value that something has *for a particular individual*. Something, such as pleasure at another's misfortune, may make life better *for that person* even though it is morally reprehensible.⁴ This notion of prudential value is inextricably entwined with the concept of well-being. Something is prudentially valuable or good for one to the extent that it contributes to one's well-being. Insofar as a theory is an account of well-being, it will also be an account of what is prudentially valuable or good.

As a doctrine about prudential value, then, hedonism holds that pleasure is the only thing that is intrinsically prudentially good—that pleasure is the only thing that is prudentially valuable in its own right and for its own sake.⁵ This version of hedonism can also be articulated in terms of well-being: well-being rests solely on the presence of pleasure. To distinguish it from the versions described above, I will label this formulation of hedonism “value hedonism.”⁶ Value hedonism is the most common philosophical form of

can nevertheless consider how well I am doing over a variable stretch of time—a day, a year, or even an entire lifetime. In formal terms, my well-being between t_1 and t_2 is equal to the integral taken over the interval from t_1 to t_2 of my well-being with respect to time.

³I borrow the term “prudential value” from James Griffin. See *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Clarendon Press, 1986), pp. 3-4.

⁴Peter Railton, “Naturalism and Prescriptivity,” *Social Philosophy and Policy* 7 (1989): 151-74, p. 155. For more on the difference between prudential and moral value, see Sumner, pp. 20-25; Railton, “Facts and Values,” *Philosophical Topics* 14 (1986): 5-31, p. 5; Thomas Scanlon, “The Moral Basis of Interpersonal Comparisons,” in Jon Elster and John E. Roemer (eds.), *Interpersonal Comparisons of Well-Being* (Cambridge: Cambridge University Press, 1991), pp. 18-20; and David Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), pp. 217-36. Some of these philosophers refer to this mode of value as “intrinsic” or “nonmoral” value. This approach to the concept of intrinsic value contrasts sharply with that of philosophers such as G.E. Moore, C.D. Broad, and Roderick Chisholm, who hold that something can be intrinsically valuable without being valuable for a particular person. For a discussion of that tradition, see Noah M. Lemos, *Intrinsic Value: Concept and Warrant* (Cambridge: Cambridge University Press, 1994).

⁵Railton, “Naturalism and Prescriptivity,” p. 155. If we assume that pain is the opposite of pleasure, then it follows that pain is the only thing that is intrinsically prudentially bad.

⁶One might be tempted to draw a distinction between hedonism as a theory of prudential value (value hedonism) and hedonism as a theory of well-being (say, welfare hedonism). As I have defined well-being and prudential value, however, it is difficult to distinguish welfare hedonism from value hedonism. They make almost identical substantive claims. I will,

hedonism; and for the remainder of this essay, when I refer without qualification to hedonism, I will be referring to hedonism as an account of well-being and intrinsic prudential value.

Hedonism has few defenders among contemporary philosophers, and it occupies only a corner of the recent literature on well-being.⁷ Philosophers generally recognize three distinct approaches to the topic of well-being: mental state or experience theories, preference or desire theories, and objective list or substantive good theories.⁸ Hedonism falls within the first category: pleasure is fundamentally experiential in nature. Thus, according to hedonism, welfare rests only on the presence of certain experiences or mental states.

Critics of hedonism are nearly unanimous in their rejection of this unrelenting "experientialism."⁹ In defense of this rejection, they appeal to a thought experiment first conjured up in 1974 by Robert Nozick: the experience machine.¹⁰ Nozick invites us to imagine a machine capable of providing us only with pleasurable experiences. This machine can stimulate our brains so that we feel we are living an ideal life. It can provide us with the experiences of writing a great novel, falling in love, or studying philosophy. And while we are on the machine, the experiences we have are indistinguishable from real experiences. If well-being consists only in psychological states or experiences, Nozick asks, why do we find the idea of a life spent on the experience machine so disturbing? Many of the most prominent philosophers

therefore, treat value hedonism as a doctrine about both prudential value and well-being.

⁷These few defenders of hedonism include Shelly Kagan, "The Limits of Well-Being," in Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul (eds.), *The Good Life and the Human Good* (Cambridge: Cambridge University Press, 1992), pp. 169-89; Rem B. Edwards, *Pleasures and Pains: A Theory of Qualitative Hedonism* (Ithaca: Cornell University Press, 1979); Fred Feldman, *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy* (Cambridge: Cambridge University Press, 1997); and Wendy Donner, *The Liberal Self: John Stuart Mill's Moral and Political Philosophy* (Ithaca: Cornell University Press, 1991).

In "Fairness to Happiness," *Social Theory and Practice* 15 (1989): 33-58, Richard Brandt defends a position that combines hedonism with a "carefully restricted form of desire theory" (p. 54). He defines the concept of a prudential good as something that is the object of a "fully and vividly informed" desire (p. 35), and then argues that happiness is the only thing that falls under that concept. According to Brandt, then, happiness is the only thing that is prudentially valuable, but it is valuable only in virtue of the fact that we desire it. For a more complete presentation of this theory, see Brandt's *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), and *Facts, Values, and Morality* (Cambridge: Cambridge University Press, 1996).

⁸Derek Parfit divides the field in this manner in *Reasons and Persons* (Oxford: Clarendon Press, 1984), p. 493, and most philosophers follow his lead with only minor adjustments. Shelly Kagan, however, calls for a different classification. See Kagan, pp. 187-89.

⁹I borrow this term from Scanlon. See Scanlon, p. 20.

¹⁰Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 42-45.

of value—including James Griffin, David Brink, Stephen Darwall, and L. W. Sumner—take this thought experiment to be the definitive response to hedonism and, more broadly, to all mental state theories of well-being.¹¹ In anthologies of moral philosophy, Nozick's experience machine is often the only argument offered in response to classical hedonism.¹² If these philosophers are correct, mental state or experientialist theories are not tenable approaches to well-being and prudential value.

As yet, no hedonistic or mental state theorist has undertaken a determined response to the experience machine.¹³ That is my project in this essay. I will

¹¹Griffin, pp. 9-10; Brink, pp. 223-24; Stephen Darwall, "Self-Interest and Self-Concern," *Social Philosophy and Policy* 14 (1997): 158-78, pp. 162, 178; and Sumner, pp. 94-98. According to Sumner, the experience machine thought experiment "strikes against any hedonistic theory" according to which "the impact on our well-being of some particular experience is entirely determined by features of the experience which are available to introspection—how it feels, how agreeable we find it, how much we wish it to continue, or whatever. . . . The lesson of the experience machine is that any theory with this implication is too interior and solipsistic to provide a descriptively adequate account of the nature of welfare" (p. 98).

J.J.C. Smart goes so far as to say that worry about Nozick's experience machine puts him "in mind of the rather similar worries which depressed J.S. Mill so much, when he wondered what point there would be in life if ever the utilitarian millennium should be achieved." "Hedonistic and Ideal Utilitarianism," in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy*, vol. 3, *Studies in Ethical Theory* (Morris, Minn.: The University of Minnesota, Morris, 1978), p. 251 n. 32. Cf. Lemos, pp. 202-3. Other philosophers who wield the experience machine against hedonism include John Finnis, *Natural Law and Natural Rights* (Oxford: Clarendon Press, 1980), p. 33, and *Fundamentals of Ethics* (Washington, D.C.: Georgetown University Press, 1983), pp. 37-42; Garrett Thomson, *Needs* (London: Routledge and Kegan Paul, 1987), p. 41; and Robin Attfield, *A Theory of Value and Obligation* (London: Croom Helm, 1987), p. 33.

¹²See Thomas L. Carson and Paul K. Moser (eds.), *Morality and the Good Life* (New York: Oxford University Press, 1997), p. 7; Louis P. Pojman (ed.), *Ethical Theory: Classical and Contemporary Readings*, 2nd ed. (Belmont, Cal.: Wadsworth Publishing Company, 1995), p. 124; Joram G. Haber (ed.), *Doing and Being: Selected Readings in Moral Philosophy* (New York: Macmillan, 1993), p. 7; and George Sher (ed.), *Moral Philosophy: Selected Readings*, 2nd ed. (Fort Worth: Harcourt Brace College Publishers, 1996), p. 612. In these anthologies, the editors explicitly oppose Nozick's thought experiment to hedonism. See also Steven M. Cahn and Joram G. Haber (eds.), *Twentieth Century Ethical Theory* (Englewood Cliffs, N.J.: Prentice-Hall, 1995).

¹³Wendy Donner claims to defend hedonism against Nozick's thought experiment (pp. 69-78). She contends that qualitative hedonism can account for the experience machine by allowing that "authentic" pleasures are qualitatively superior to the "hallucinatory" pleasures supplied by the machine. This approach challenges the very premise of the experience machine argument by supposing that the machine-produced experiences are not actually identical to their "authentic" counterparts (p. 78). In particular, Donner suggests that an experience is defined by both its subjective "feel" and its source or cause; experiences can be individuated by reference to external or objective facts (p. 77). Thus, although the experiences caused by the machine are subjectively indistinguishable from their authentic counterparts, they are, according to Donner, nonetheless qualitatively different experiences.

take a close look at the thought experiment and the ways Nozick and others use it, and I will attempt to demonstrate that it does not constitute a refutation of mental state accounts of prudential value. This demonstration will not focus entirely on the narrower doctrine of hedonism and its correspondent notion of pleasure. The nature of pleasure is itself quite controversial, and a mental state theorist need not consider pleasure the only intrinsically good experience. Following both Nozick and Railton, I will focus instead on hedonism and experientialism in terms of happiness, where happiness is viewed as a psychological state or class of such states.¹⁴ I am, then, assuming an experientialist account of happiness. Anyone inclined to reject such an account, however, need not jump ship prior to departure. This assumption is a semantic decision more than anything else; I could just as well forswear the mental state definition of happiness and opt for a more psychologically-oriented term such as "satisfaction" or "enjoyment."¹⁵ For the sake of brevity and familiarity, though, I will use "happiness."

Nozick's Experience Machine

Let us turn now to the details of the thought experiment and the argument as Nozick presents them. To demonstrate that we care about more than how our lives feel to us from the inside, Nozick asks us to consider the following scenario:

Imagine a machine that could give you any experience (or sequence of experiences) you might desire. When connected to this experience machine, you can have the experience of

She can, therefore, agree that a life on the experience machine is not the best life without appealing to anything outside of experience. See also Tyler Burge, "Other Bodies," in Andrew Woodfield (ed.), *Thought and Object: Essays on Intentionality*, (Oxford: Clarendon Press, 1982), pp. 97-120. I do not accept this account of experience, and so I do not consider the theory of well-being that it produces to be an experientialist one. Donner manages to sidestep Nozick's argument only by taking the bite out of experientialism. By connecting experience to external fact, she blurs the distinction on which hedonism is based. In this essay, I am interested in defending the more radical position that is concerned only with the subjective content of experience. We could, however, accept Donner's definition of experience and simply reformulate hedonism in terms of that more narrow, subjective content.

¹⁴Railton, "Naturalism and Prescriptivity," p. 154. Railton says, "I use the expression 'happiness' rather than 'pleasure' in characterizing hedonism because of the unfortunately narrow connotations of the latter. I do intend 'happiness' to pick out a class of experiential states" (p. 154 n. 9). This shift from "pleasure" to "happiness" is not necessary. Most hedonists hold a pluralistic conception of pleasure, according to which pleasure is not a single experience or felt quality of experience, but rather a class of diverse experiences that have in common merely the fact that they are liked or desired. The everyday sensory connotations of "pleasure" remain misleading, though; thus "happiness" is my term of choice.

¹⁵See Darwall, p. 161.

writing a great poem or bringing about world peace or loving someone and being loved in return. You can experience the felt pleasures of these things, how they "feel from the inside." You can program your experiences for tomorrow, or this week, or this year, or even for the rest of your life. If your imagination is impoverished, you can use the library of suggestions extracted from biographies and enhanced by novelists and psychologists. You can live your fondest dreams "from the inside." . . . Upon entering, you will not remember having done this; so no pleasures will get ruined by realizing they are machine-produced.¹⁶

Nozick then asks us whether we would choose to connect ourselves to such a machine for the rest of our lives. He contends that the majority of us would not. Our unwillingness to surrender to the experience machine, according to Nozick, betrays the fact that "we care about things in addition to how our lives *feel* to us from the inside."¹⁷ It proves that we want more than pleasurable experiences. We want certain states of affairs actually to obtain. And so, Nozick concludes, "there is more to life than feeling happy."¹⁸

Before I begin my own analysis of Nozick's argument, I want to consider two possible but ultimately unconvincing objections to the details of the thought experiment itself. These objections purport to show that the thought experiment fails to isolate the intuitions and desires that Nozick claims it does. The first involves the suggestion that our refusal to plug in is only the result of insecurities on our part regarding the machine's proper functioning. Sumner mentions a few such potential anxieties. "How do we know that the technology is foolproof? What happens if there is a power failure? Suppose the operators of the machine are really sadistic thrill-seekers, or the premises are overrun by fundamentalist zealots."¹⁹ What if we grow bored with pure pleasure or happiness? These worries only call for us to be diligent in the use of our imagination. If we can imagine the technology itself, we can no doubt imagine its reliability and efficiency as well. Moreover, we might allay the more serious fears by altering Nozick's scenario a bit. For example, suppose we are allowed to sample the experience machine before committing our lives to it. We will know, then, whether it actually works. Suppose further that the machine will throw in just enough misery and pain to add the requisite flavor to our other experiences.²⁰ If we exercise a bit of mental dexterity,

¹⁶Robert Nozick, *The Examined Life: Philosophical Meditations* (New York: Simon and Schuster, 1989), pp. 104-5. Jonathan Glover offers a veritable plethora of alternative versions of the experience machine in his *What Sort of People Should There Be?* (New York: Penguin Books, 1984), pp. 91-113. For a fascinating comparison, see a thought experiment that predates Nozick's by only a year in J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), pp. 12-27.

¹⁷Nozick, *The Examined Life*, p. 104.

¹⁸*Ibid.*, p. 106.

¹⁹Sumner, p. 95.

²⁰Jeffrey Goldsworthy, in "Well-Being and Value," *Utilitas* 4 (1992): 1-26, has less faith in our imaginations. He contends that even these reassurances do not constitute an accept-

therefore, we can make an earnest and largely successful effort to overcome the sorts of doubts raised by Sumner. Even without these doubts, though, most of us continue to share Nozick's intuitions: we remain unwilling to accept a lifetime on the experience machine.

The second unconvincing objection involves those few moments between our decision to connect and the flipping of the switch. Perhaps our rejection of the experience machine stems only from the horror we would feel during that interval. If that is so, then our refusal to connect is motivated by the fear of psychological torment, and therefore the thought experiment fails to isolate intuitions or wants regarding anything other than how we *feel*. Nozick anticipates this objection. You would not refrain, he maintains, "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?"²¹ If in fact we care only about our experiences, a few moments of displeasure would not—assuming we are rational—move us to throw away a life full of happiness. Moreover, to the extent that we acknowledge we would feel distress, we admit to possessing exactly the intuitions to which Nozick is pointing. We want more than a life of artificial experiences, and that is why we would feel anxiety before connecting.²²

Nozick's argument does succeed, then, in isolating the fact that we care about more than our experiences. It then concludes that there is more to well-being than how our lives feel to us from the inside. But it is unclear how one arrives at this claim merely from the fact that we care about more than happiness. One might be misled by a certain ambiguity in Nozick's language. In his discussion of the experience machine, Nozick often moves back and forth between asking whether "we care about things" in addition to our internal feelings and asking whether "only our internal feelings matter to us."²³ Nozick's argument trades on the ambiguity of the crucial term "matter." That something matters can mean merely that we desire it and care

able guarantee. "It seems impossible that *any* purported proof or guarantee could make the risk [of connecting to the machine] worthwhile" (p. 18). See Glover, pp. 93-94, for a more thorough response to this "primitive" objection.

²¹Nozick, *Anarchy, State, and Utopia*, p. 43.

²²A third possible objection turns on a metaphysical question. Do the machine experiences themselves constitute a real world? In "Fairness to Happiness," Brandt suggests that the world of the experience machine is no different from Berkeley's world, in which we are merely souls not even existing in real space whose experiences are supplied by God (p. 49). If the dream world of the experience machine is as real as the "real world," then the notion of how our lives feel from the *inside* as opposed to the outside makes no sense. While I find this a fascinating possibility, I assume in this essay that there is, in fact, a real world, and that the world of the experience machine is in some way distinguishable (perhaps causally). Glover's discussion of this metaphysical problem is again very illuminating (pp. 105-7).

²³Nozick, *The Examined Life*, pp. 104, 105.

about it, but it can also imply that it is prudentially valuable. If we say, for instance, that it *matters* to me whether it rains today, we are probably making the weaker claim that I have hopes and desires whose satisfaction depends upon the weather. But if we say that philosophy is all that *matters* in life, we may very well be making the stronger assertion that philosophy is all that is intrinsically prudentially valuable, that our well-being depends upon our engagement in some sort of philosophical activity.

This term “matters” occupies a central place in Nozick’s argument, which might be summarized as follows.

(1) “We care about things in addition to how our lives *feel* from the inside.”²⁴

(2) It is not the case that “all that *matters* about a life is how it feels from the inside.”²⁵

(3) “There is more to [a good] life than feeling happy.”²⁶ Happiness is not the only thing that is intrinsically prudentially valuable; well-being does not consist of happiness alone.

Because of the ambiguity of “matters,” this argument may appear relatively uncontroversial. The assertion that we care about something is synonymous with the assertion that something matters to us—given the first sense of the term “matters”—and so (2) follows from (1). Given the other sense of “matters,” if something other than happiness matters to our lives, then there is more to well-being than merely happiness, and hence (3) follows from (2). It appears, therefore, that (1) does in fact imply (3). This implication relies, however, upon a slippery exchange of terms that merely look the same. “Matters” in the move from (1) to (2) does not mean the same as “matters” in the move from (2) to (3), and to assume that what matters in the first sense also matters in the second is to beg the question at hand. This connection between the two senses of the word “matters”—what we care about versus what is central to our welfare—is precisely what the critic of hedonism must demonstrate. Without the addition of further premises and arguments, Nozick’s argument is invalid, and it thereby fails to refute hedonism.

Wants and Well-Being

The experience machine argument establishes first that we care about more than machine-produced happiness. This is the premise that is then supposed

²⁴Ibid., p. 104; emphasis in original.

²⁵Ibid., p. 105; emphasis added.

²⁶Ibid., p. 106.

to yield the conclusion that there is more to well-being than happiness. While it is unclear just what this premise involves, most opponents of hedonism who appeal to the experience machine concentrate on the *wants* or *desires* that the thought experiment isolates. In particular, they focus on our desire to “track reality”—our desire for an “actual connection with reality.”²⁷ The additional premises and arguments required for the experience machine to refute hedonism, then, must involve some sort of connection between our desires and our well-being. The hypothesis of such a connection is not at all unreasonable. Any determination of what there is to the good life must rely on, or at least begin with, our wants and our cares. How else can we begin to identify the good life except by looking at those lives we in fact want, those lives we intuitively hold to be good?

The nature of this connection between our wants and our welfare is unclear. If the connection is not secure, then the experience machine argument against hedonism leaves open the possibility that, while in general our desires are closely related to our well-being, the ties unravel in the case of the thought experiment. In this section, I will consider the various ways in which a supporter of Nozick’s argument might fill in the details of this relationship between our desires and our well-being.

The first and simplest approach to this relationship is to contend that the satisfaction of our desires is constitutive of our well-being. This view—the actual-desire-satisfaction account of well-being—equates our welfare with the satisfaction of our actual or revealed desires. We desire to avoid the experience machine, and so there must be more to our well-being than the happiness that the machine can provide. The philosophers who defend the experience machine argument are, however, as critical of the actual-desire-satisfaction account of well-being as they are of hedonism. Griffin correctly observes that “the objection to the actual-desire account is overwhelming. . . . Notoriously, we mistake our own interests. It is depressingly common that when even some of our strongest and most central desires are fulfilled, we are no better, even worse, off.”²⁸ The connection between our desires and well-being must therefore be more complicated than that posited by the actual-desire theory.

The rejection of the actual-desire-satisfaction approach to the gap between our desires and our well-being suggests a more complicated alternative. Instead of identifying the satisfaction of desires with welfare, this second approach views our desires as conclusive evidence, as a guide to a good life. We identify a good life by examining what it is that we desire. Not

²⁷Ibid., p. 106. See also Sumner, p. 96, and Griffin, p. 9.

²⁸Griffin, p. 10. For further criticisms of and comments about actual-desire-satisfaction theories, see Sumner, pp. 118-22; Donner, pp. 79-82; Parfit, p. 494; and Brandt, *A Theory of the Good and the Right*, pp. 247-48.

just any desire will do. We are searching for an account of *human* well-being, and so we must rely on those desires that are nearly universal in scope. (No preferences are entirely universal.) These near-universal desires, then, point to the things that are intrinsically prudentially valuable. This approach provides the connection needed by the experience machine argument. Nozick's thought experiment establishes that when we are confronted with the possibility of the experience machine, we have a near-universal desire to avoid it, to remain connected to reality. If this new explanation of the relation between desires and well-being is correct, then the experience machine thought experiment poses a strong challenge to hedonism.

Our near-universal desires do not, however, serve as conclusive and foolproof evidence for what constitutes our well-being. Although our cares and desires may provide hints, they do not by themselves point to our well-being. Consider the near-universal desire for wealth. In our society wealth often contributes to well-being, but it is not a necessary element of a good life. There are people in the world who live good lives without wealth. They may—along with the rest of society—desire to be wealthy, but they do not lack well-being merely because the state of affairs that is desired is never actualized. We cannot, therefore, blindly appeal to what we all care about in order to determine what is and is not intrinsically prudentially valuable. As indicators of well-being, our desires are fallible.

This fallibility must be accounted for in order to preserve the strong connection between our desires and our welfare, a connection that must be preserved if the experience machine argument is to refute hedonism. One option is to introduce a distinction that explains which of our desires are generally reliable indicators of value. In particular, intrinsic desires—those things that we want and about which we care *for themselves*—can be distinguished from instrumental desires. Thus, while not every near-universal desire is a reliable indicator of well-being, we might be able to appeal safely to our near-universal *intrinsic* desires. Our desire for wealth is *not* an intrinsic desire: we want wealth not for itself but rather because we believe that it will aid us in the acquisition of things we do value for themselves.

This appeal to near-universal intrinsic desires does not ultimately vindicate the desires-as-evidence approach. Consider the following thought experiment. Imagine a world in which the evil tobacco companies have adulterated their cigarettes with a chemical that makes them enormously psychologically addictive. Before long, most of the world is hooked on cigarettes, and we all feel a constant and tremendous desire to smoke. We care about smoking so much that we tend to neglect other aspects of our lives, and we are unhappy as a result. We do not understand our craving, nor do we derive any pleasure or enjoyment from its satisfaction. All we end up with is the desire for another smoke. Moreover, this desire for cigarettes is

a near-universal intrinsic desire; we all care about smoking *for itself*. When the soon-to-vanish American Cancer Society presents us with a proven cure that will alleviate our desire to smoke and allow us to pursue other things such as our own happiness, we refuse. We do not care about happiness or other ends. All we want is another drag, and we want it for nothing but itself.

According to the modified desires-as-evidence approach, in this imaginary world smoking is the only thing that is prudentially valuable. This thought experiment posits a world quite unlike our own. In our world, the satisfaction of a desire almost always involves a sense of pleasure or happiness.²⁹ In the world of this thought experiment, though, the satisfaction of our desires does not involve any pleasure or happiness whatsoever.³⁰ We have only one desire—the desire to smoke—and the satisfaction of that desire brings us no enjoyment at all. As a result, it may be difficult to regard the state of affairs that satisfies that desire as prudentially valuable. Our intuitions about the direct connection between prudential value and our desires seem to weaken when we consider a world in which the satisfaction of those desires never involves a concomitant experience of happiness. Even our near-universal intrinsic desires are, therefore, at best fallible indicators of prudential value, and we cannot rely on them to provide conclusive evidence about our well-being. They too fail to bridge the gap between our wants and our welfare.

This failure constitutes a general objection to the experience machine argument against hedonism. If desires for things for themselves do not always indicate that such things are prudentially valuable, then the fact that we care about more than happiness does not demonstrate that there is more to a good life than happiness. That is, there are no viable premises that can be added to Nozick's argument in order to make it valid. Nonetheless, the experience machine argument is not dead. Its defenders can still appeal to the fact that the search for our well-being must at least *begin* with our cares and wants. Although our desires and intuitions are not always accurate, we should not disregard them entirely. Surely, defenders maintain, our overwhelming intuitive prejudice against the experience machine tells us *some-*

²⁹The term "satisfaction" may be misleading here. When I say that a desire is satisfied, I mean that the state of affairs that is the object of that desire obtains, and that the person whose desire it is recognizes that it obtains.

³⁰One might object that such a world is a psychological impossibility—that it is a truth of human beings that the satisfaction of a desire entails a pleasant experience. Alternatively, one might simply *define* happiness and the satisfaction of desires in terms of one another and argue that the imaginary world of the thought experiment is a *necessary* impossibility (see n. 14). But if happiness is intertwined (either psychologically or necessarily) with desire-satisfaction, then the distinction between hedonistic and desire-satisfaction accounts of well-being begins to fade. This is a fascinating possibility, but I do not have the space here to discuss it at length.

thing about the role of happiness in a good life.

Hedonism is *prima facie* an intuitive theory. It directly captures our most central intuition about well-being: if anything is good for a person, the experience of happiness is. Hedonism can also capture most of our other intuitive judgments. Almost all of the things we intuitively regard to be prudentially valuable are also things that make us happy, and so these intuitions fit comfortably within the hedonistic scheme. Nozick's thought experiment isolates intuitions that hedonism cannot directly capture. It foregrounds our intuitive views about a life spent on the experience machine. So far I have been attempting to lessen the force of those intuitions by arguing that they are fallible. The theory we abstract from them—namely, that there is a direct and absolute connection between our wants and our welfare—is incorrect. I believe we can lessen the force of those experience machine intuitions even more. In the next section, following a strategy suggested by Peter Railton in his article "Naturalism and Prescriptivity," I will begin to explore ways in which we can explain away the intuitions that hedonism cannot directly capture.³¹

Other Values

One way to lessen the force of the experience machine intuitions is to demonstrate that they are really about something other than well-being. There are several different dimensions of value, and it is often difficult to determine from which dimension our intuitions arise. Of the many types of value that might be lost in a life on the experience machine, perhaps the most obvious is moral value. Moral value involves the impact our lives and actions have on others. Choosing to connect to the experience machine seems to many people to be a morally repugnant decision. Such a decision involves a complete abandonment of any persons, relationships, or activities in the real world. When you are on the machine, you cannot contribute to *real* political causes, you cannot repay any *real* debts, and you cannot help *real* people in need. In this case our moral intuitions are clear: a life on the experience machine seems morally bankrupt. It is therefore possible that our rejection of the experience machine is driven by intuitions regarding moral value, not prudential value. Our intuitions *do* suggest that something of value is lost when we connect, but it is not clear that what we lose is well-being.

Nozick anticipates this objection, and he modifies his scenario to ensure

³¹Although Railton presents this viewpoint and considers its strengths, he does not advocate it. He is not a hedonist; his own position is a form of the desire-satisfaction theory. Railton does believe, though, that this line of argument is a strong one, and I will therefore continue to refer to it as his.

that our rejection of the machine is not due to moral concerns: "Other people also have the same option of using these machines which, let us suppose, are provided by friendly and trustworthy beings from another galaxy, so you need not refuse connecting in order to help others."³² Nozick argues persuasively that even with this modification we do not want to connect to the machine. Our persistent refusal cannot be due to moral considerations. If everyone is on the machine, our moral obligation to help others is irrelevant. If hedonism is correct, then there is nothing we can do for them that the machines do not provide. By modifying the thought experiment, then, Nozick successfully dispels any confusion between moral and prudential value judgments.

This maneuver does not eliminate all possible confusion, however. There are other dimensions of value that are not so easily isolated. Consider the realm of aesthetics. We attribute aesthetic value to those things we find attractive or admirable; and at any particular moment, a human life may be more or less aesthetically valuable.³³ When we entertain Nozick's thought experiment, we find life on the machine unattractive: it is aesthetically displeasing. Perfectionism is another dimension of value. We ascribe perfectionist value to those things that exemplify or excel in the characteristics of their nature.³⁴ A human life has perfectionist value to the extent that it develops or realizes what is central to human nature. On most plausible accounts of human nature, perfectionism involves the pursuit of activities in the real world. According to Marx, for example, humans are both productive and social; the most perfect life, then, is one that includes both labor and social interaction.³⁵ On the experience machine, though, one cannot engage in the activities that are essential to our nature.³⁶ Life on the experience machine lacks both aesthetic and perfectionist value.

Our consideration of the experience machine thought experiment in the

³²Nozick, *The Examined Life*, p. 105.

³³For a more complete description of aesthetic standards for human lives, see Sumner, pp. 21-23.

³⁴Thomas Hurka provides an expansive account of this dimension of value in *Perfectionism* (New York: Oxford University Press, 1993). See also Sumner, pp. 23-24.

³⁵See Marx's *Economic and Philosophical Manuscripts* (1844), in Karl Marx, *Early Writings*, trans. Rodney Livingstone and Gregor Benton (New York: Vintage Books, 1975). Marx writes: "The worker can create nothing without *nature*, without the *sensuous external world*" (p. 325). "It is therefore in his fashioning of the objective [world] that man really proves himself to be a *species-being*. Such production is his active species-life" (p. 329).

³⁶In "Excellence: Trying, Deserving, Succeeding," in Dudley Knowles and John Skorupski (eds.), *Virtue and Taste: Essays on Politics, Ethics and Aesthetics* (Oxford: Blackwell, 1993), Thomas Hurka notes that according to some accounts of human nature, life on the experience machine does not lack perfectionist value (p. 60). Aristotle's view of human nature as rational activity might—on a certain understanding of rationality—count as one such account. It is clear, though, that Hurka prefers a more robust version of perfectionism.

context of prudential value is clouded by our concern for these other modes of value. Railton notes that "a distinction must be made between those elements of what is important to me that depend upon their contribution to how well my life goes and those elements that have to do with other forms of value . . . my life might have for myself or others."³⁷ However, our initial reaction to the machine is not informed by this theoretical distinction. Taken at face value, our intuitions cannot serve as evidence against value hedonism. We use a variety of standards to evaluate our lives, and our rejection of the experience machine might be due to any of them. We may have mistaken what is intrinsically prudentially valuable for what is valuable in some other sense.

James Griffin disagrees. He maintains that he prefers to track reality "not because it would be morally better, or aesthetically better, or more noble, but because it would make for a better life for me to live."³⁸ Griffin does not offer an argument in support of this contention, and the difficulty involved in isolating intuitions arising from a particular sphere of value should make us wary of his confidence. Our lives would certainly lose various sorts of value were we connected to the experience machine. We cannot assume, however, as Griffin does, that the machine detracts from our *well-being*. Nor can we alter the thought experiment to rule out aesthetic or perfectionist concerns. Although the different dimensions of value are conceptually distinct from one another, our intuitions about well-being are intertwined with our intuitions about aesthetics and perfectionism; and in this case, we might simply be unable to disentangle them. There remains, therefore, a possibility that the intuitive views to which the experience machine argument appeals are not, in fact, views about well-being. This continuing uncertainty lessens the force of those intuitions.³⁹

The consideration of other modes of value yields more than just this uncertainty, though. Meditation upon the distinctions between these different types of value might lead us to revise our initial conviction that the experience machine is detrimental to our welfare. When we return to the thought experiment with these distinctions firmly in mind, the possibility that a life on the machine is full of prudential value might seem less objectionable. A life on the machine, while it might be unworthy of a human being and thus lack perfectionist value, might nonetheless be going quite well for the person who is living it. If this is so, Sumner observes, "then the lesson of the experience machine may be, not that mental state theories are deficient as accounts of the nature of welfare, but that welfare tracks only one dimension

³⁷Railton, "Naturalism and Prescriptivity," p. 170.

³⁸Griffin, p. 9.

³⁹For a similar argument, but one in terms of what is good for us versus what is good *simpliciter*, see Goldsworthy, pp. 18-20.

of the value of a life."⁴⁰

The recognition of other values lessens, but does not eliminate, the force of our experience machine intuitions. We still seem to have strong intuitive views about what is prudentially valuable, even if we cannot always distinguish those views from beliefs about other modes of value. Much more can be done to lessen the force of those intuitions. As Railton observes, if our diverse intuitions about what is prudentially valuable "can be explained in a unified way by invoking a substantive conception of happiness . . . then the hedonist can claim that, despite appearances, these other ends owe their hold upon us to the role they have played in the creation of happiness."⁴¹ In the next two sections I will consider and defend just such an explanation.

Happiness and Desire Conditioning

The first component of this explanation is a psychological account of the formation of our desires. Richard Brandt suggests that happiness is the natural source of all our desires and aversions. When we associate pleasure or happiness with a certain experience or state of affairs, that experience or state of affairs becomes the object of a desire. According to Brandt, this process of conditioning is the *only* fundamental process involved in the acquisition of desires.⁴² Railton outlines a similar model.

When a given set of desires leads us to act in a way that brings with it the attainment of happiness, these desires are positively reinforced; conversely, when other desires lead us to act in ways that lead to unhappiness, they are negatively reinforced. Over time for any given individual and relative to the range of behaviors he undertakes, individuals will tend to possess and act on desires that have brought happiness in the past. Quite likely, most of these desires will have immediate objects other than happiness, and many will involve intrinsic interest in ends other than happiness. What evolves in the individual, then, is a set of desires, including intrinsic desires, that can be explained in part as tracing a path oriented toward the experience of happiness, even though individuals often do not aim at happiness.⁴³

⁴⁰Sumner, p. 96.

⁴¹Railton, "Naturalism and Prescriptivity," p. 169.

⁴²Brandt, *A Theory of the Good and the Right*, p. 100. This seems to me to be the most controversial claim that Brandt makes on behalf of this psychological theory. My arguments do not depend, however, on happiness being the *only* source of our desires. Even if it is not the whole story, it is certainly a large part of the story, and that is enough to support my arguments that follow. Note that Brandt's theory does not affirm that we desire *only* happiness. The fact that past experiences of happiness determine our present desires does not, Brandt observes, imply that our present desires are only for experiences of happiness. "What we want is an event of a certain sort, which could be getting a piece of knowledge or enhancement of the welfare of our children" (*Facts, Values, and Morality*, p. 29).

⁴³Railton, "Naturalism and Prescriptivity," p. 167.

This model explains how we come to desire things other than our own happiness, and how we begin to desire those things *for themselves*, independently of the happiness they produce. Initially, for example, I may desire to study philosophy because it gives me great pleasure and happiness. Those experiences of happiness reinforce that desire through the extended process of conditioning. Eventually, my desire to philosophize takes on a life of its own. Although it continues to give me happiness, I begin to desire to study philosophy for its own sake.⁴⁴ I would continue to study philosophy even if I found that it no longer made me happy to do so.⁴⁵ It is, according to Brandt, "just a fact of human nature that we learn to like (want) for themselves things which reliably lead to other things we already like (want)."⁴⁶ Our desires have a tendency to outstrip their happiness-based origins; and ironically, it is the fact that certain events are accompanied by experiences of happiness that ultimately propels us to desire things other than happiness.⁴⁷

The second component of the explanation for our diverse intuitions is found in the late nineteenth-century writings of Henry Sidgwick. In the fourth chapter of *The Methods of Ethics* entitled "Pleasure and Desire," Sidgwick describes the "fundamental paradox of Hedonism, that the impulse towards pleasure, if too predominant, defeats its own aim."⁴⁸ According to

⁴⁴In the section of *Utilitarianism*, ed. George Sher (Indianapolis: Hackett Publishing Company, 1979), entitled "Of What Sort of Proof the Principle of Utility is Susceptible," John Stuart Mill provides a similar account of the desire for money. "What, for example, shall we say of the love of money? There is nothing originally more desirable about money than about any heap of glittering pebbles. Its worth is solely that of the things which it will buy; the desires for other things than itself, which it is a means of gratifying. Yet the love of money is not only one of the strongest moving forces of human life, but money is, in many cases, desired in and for itself; the desire to possess it is often stronger than the desire to use it, and goes on increasing when all the desires which point to ends beyond it, to be compassed by it, are falling off" (p. 36). In *Facts, Values, and Morality*, Brandt makes a parallel claim regarding our desire for knowledge. "Why might not . . . reinforcement by pleasure produce a strong desire for knowledge in me, a desire so strong that I regularly set aside pleasure in order to gain more knowledge (p. 193).

⁴⁵I would not continue my philosophical studies indefinitely. Repeated unpleasant experiences would initiate a process of counter-conditioning, and eventually I would no longer desire to philosophize at all.

⁴⁶Brandt, *A Theory of the Good and the Right*, p. 122.

⁴⁷Brandt provides a thorough and persuasive defense of this theory as well as a summary of the psychological evidence in *A Theory of the Good and the Right*, pp. 88-103, and *Facts, Values, and Morality*, pp. 21-33. Ultimately, though, this theory falls within the domain of psychology; and as Railton observes, we should be content to leave its assessment largely to the development of psychological theory. See "Naturalism and Prescriptivity," p. 167.

⁴⁸Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1913), p. 48. John Stuart Mill also defends this idea. In his *Autobiography*, ed. Jack Stillinger (Boston: Houghton Mifflin Company, 1969), Mill writes, "I now thought that this end [happiness] was only to be attained by not making it the direct end. Those only are happy (I thought) who have their minds fixed on some object other than their own happiness. . . . Aiming thus

Sidgwick, we cannot attain the greatest scope of pleasure and happiness if our cares and desires take only pleasure and happiness as their ends. In order for us to achieve the highest happiness, our desire for happiness must be superseded by "other more objective, 'extra-regarding,' impulses."⁴⁹ A quick observation lends credence to this claim. When we look out at the world, we find that more often than not, the people who are most happy are those who have spent their lives in pursuit of a plurality of goods. The "paradox," then, is that the most effective way to be happy is to care more about things other than happiness than about happiness itself.⁵⁰

Together these two components form a psychological account of the relationship between happiness and desires. Brandt and Railton defend a theory according to which all of our desires are conditioned by experiences of happiness and unhappiness. Let us now consider that theory in light of Sidgwick's "paradox." According to Sidgwick, whenever we exclusively pursue our own happiness, that happiness invariably eludes us. Thus the desire to seek only happiness is constantly eroded by the process of conditioning. Meanwhile, our desires for things other than happiness—for things that tend to promote happiness, such as friendship and autonomy—are gradually reinforced. As the result of this complex process, therefore, we have been conditioned not only to desire things other than happiness, but to desire those things more than happiness itself.

This account of happiness and desire conditioning suggests a particular picture of human psychology and motivation, a picture in which all of our desires stand in a determinate relationship to happiness. All of our desires, even those for things other than happiness, owe their existence to experiences of happiness. Our desires to pursue ends such as friendship and autonomy are continually reinforced by the lasting happiness those ends bring. Likewise, any desire to pursue happiness directly and exclusively is negatively reinforced by the misery that inevitably results from such pursuit. In a sense, then, happiness is psychologically prior to our desires: we must appeal to different experiences of happiness and unhappiness to explain *why* we desire the things that we do. And by employing just such an explanation,

at something else, they find happiness by the way" (pp. 85-86).

⁴⁹Sidgwick, p. 49. Sidgwick supplies numerous examples of this "paradox." "The pleasures of thought and study can only be enjoyed in the highest degree by those who have an ardour of curiosity which carries the mind temporarily away from self and its sensations. In all kinds of Art, again, the exercise of the creative faculty is attended by intense and exquisite pleasures: but it would seem that in order to get them, one must forget them: the genuine artist at work seems to have a predominant and temporarily absorbing desire for the realisation of his ideal of beauty" (p. 49).

⁵⁰For an extended and illuminating discussion of the paradox of hedonism, see Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13 (1984): 134-71, pp. 140-46.

we can further weaken those experience machine intuitions that hedonism cannot directly capture.

The Machine Breaks Down

The experience machine thought experiment appeals to our intuitions as evidence against hedonism. Our intuitions tend to reflect our desires and preferences. In particular, our experience machine intuitions reflect our desire to remain connected to the real world, to track reality. But according to the account of the relation between happiness and our desires outlined above, the desire to track reality owes its hold upon us to the role it has played in the creation of happiness. We acquire our powerful attachment to reality after finding again and again that deception almost always ends in suffering. We develop a desire to track reality because, in almost all cases, the connection to reality is conducive to 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. The experience machine argument appeals to our desires and motivations; but as Railton warns, we must "not be captivated by the surface diversity of our intuitive notions about good."⁵¹ We must not ignore the system that underlies the creation of those desires. Our desire to track reality—like all of our intrinsic desires—is related to happiness in an important way: it owes its existence to happiness. Even though it leads us away from happiness in the case of the experience machine, our desire to track reality points indirectly to happiness. And according to Sidgwick's "paradox," so does our refusal to pursue our own happiness directly by connecting to the experience machine. Any inclination to seek happiness directly is thus overwhelmed by our desire to track reality. Happiness itself, therefore, is what fundamentally effects our intuitive fear of the experience machine, and we must not be misled by that fear. 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. Our experience machine intuitions do not disappear, but they now fit comfortably into a hedonistic theory of well-being.

The genetic and evolutionary nature of the relationship between happiness and desires suggests an even stronger claim, though. The experience machine argument assumes that our desires are in some way linked to our

⁵¹Railton, "Naturalism and Prescriptivity," p. 169.

well-being. However, if we are to appeal to our desires in the search for prudential value, we must understand the nature of the system of which those desires are a part. To “take our theoretically unexamined intuitions at face value”—to rely indiscriminately on our desires as guides to our well-being—is, Railton contends, “to misunderstand the character of our own motivational system.”⁵² Happiness stands at the center of our motivational system; it is the ultimate source of our desires. Thus, all of the evidence to which the experience machine argument appeals—all of our desires, including our desire to avoid the machine itself—ultimately points towards happiness as the source of prudential value. But that is precisely the doctrine of hedonism.

Hedonism explains what would otherwise appear to be a mysterious coincidence, namely, the fact that all of our desires point towards happiness. No non-hedonistic theory of well-being can account for this fact. In other words, without 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. So the psychological account of happiness and desires does more than account for the experience machine intuitions. It actually suggests that hedonism is true. In Railton’s words, “our initial confidence that ends other than happiness figure in their own right in a person’s good might in this way be undermined by reflection upon how those other ends came to seem desirable to us.”

This response to the experience machine argument does not impugn the philosophical value of all of our intuitions and desires. We desire the things that tend to make us happy. Our preferences remain, therefore, a valuable tool; they can help us identify the sources of our happiness, and hence of our well-being. We can now explain, however, *why* these usually reliable intuitions fail us so severely in the case of the experience machine. We develop a desire to track reality in a world in which detachment from reality is painful. The thought experiment marks a radical departure from this world and the circumstances under which this desire was formed. Moreover, the “paradox” of hedonism teaches us that when we pursue our own happiness, it invariably eludes us. The experience machine scenario is an exception to this teaching; it is one of the rare situations in which if we think of *only* our own happiness, that is exactly what we will attain. We are unprepared, however, to respond to the machine in this way. We have been programmed, as it were, to recoil in horror from such a departure from reality, and we have been conditioned to aim for ends other than our own happiness. Our intuitions and desires have been framed in response to a world that is substantively different from the world of the experience machine. This thought

⁵²Ibid.

experiment cannot, therefore, "afford a crucial test against experiential conceptions of the good, because it draws upon intuitions about what we want for its own sake which were developed in settings where the drastic split the machine effects between experience and reality does not typically exist."⁵³ The experience machine argument rests on an appeal to our ordinary intuitions; and as R.M. Hare observes, "these are designed to deal with ordinary cases. They give no reliable guide to what we ought to say in highly unusual cases."⁵⁴

The fact that we all intuitively reject the experience machine is merely a sign that our intuitions are functioning properly, that we are prepared to find happiness in the real world, where the failure to track reality inevitably has painful consequences. It is, then, a good thing that we desire to remain connected to reality, even though our lives would be better for us if we connected to the experience machine. Is this paradoxical? Not, says Hare, "to anybody who understands the realities of the human situation. What resolves the paradox is that the example *is* imaginary and that therefore people are not going to have to pronounce, as a practical issue," on the prudential value of a life on the experience machine.⁵⁵

Conclusion

My response to the experience machine has progressed through several distinct arguments. Some philosophers regard the experience machine thought experiment as a direct refutation of hedonism and experientialism. In the first several sections of this essay, I argued that such a refutation requires an infallible connection between the things that we desire and the things that are desirable. The direct argument against hedonism depends upon our ability to appeal to desires and intuitions in order to locate our well-being; it must bridge the philosophical gap between our wants and our welfare. Our desires and our well-being are indeed fundamentally connected: there would be no value in the world were there not entities with values, desires, and interests. I have demonstrated, however, that this connection is fallible. The fact that we intrinsically desire things other than happiness does not prove that those things are intrinsically prudentially valuable. Our desires and values are imperfect indicators of our well-being; and, therefore, the connection between our desires and our well-being is not strong enough to

⁵³Ibid., p. 170.

⁵⁴R.M. Hare, "What is Wrong with Slavery," *Philosophy and Public Affairs* 8 (1979): 103-21, p. 110; see also his discussions of intuitions and critical reasoning in *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981), pp. 47-51, 131-42.

⁵⁵Hare, "What is Wrong with Slavery," p. 115.

support the direct argument against hedonism.

Most philosophers recognize this fallibility and take the experience machine argument to be an *indirect* refutation of hedonism. 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 must provide some sort of information about our well-being.

I have offered three arguments against this appeal to our intuitions. The first of these is my contention that our rejection of the experience machine involves more than our intuitions regarding prudential value. Well-being is only one of the many standards we use to assess our lives. When we evaluate our options, there are always moral, aesthetic, and perfectionist considerations, and our intuitions regarding the experience machine may not involve considerations of prudential value at all. When we contemplate the thought experiment more carefully, we may find that while a life spent connected to the machine is bad for many reasons, it is nonetheless an intrinsically good life *for the individual on the machine*.

My second argument seeks to lessen the force of our intuitions even further. It focuses on the evolutionary and genetic relationship between happiness and desires. The experience machine argument appeals to intuitions about our desires and motivations. As Brandt and Railton see it, however, happiness is at the center of our motivational system; it drives the formation of all of our desires. Even our desire to avoid the experience machine owes its existence to past experiences of happiness. The experience machine argument misses the mark, therefore, if it focuses only on our desires and intuitions. When we recognize that those intuitions ultimately point towards happiness, the force that they bring to bear against hedonism dwindles.

This second argument suggests another—even stronger—conclusion: any appeal to our desires in the search for prudential value will ultimately gesture towards the truth of hedonism. The experience machine argument makes just such an appeal, but it takes our desires and intuitions at face value. We must also, however, attend to the nature of the system that underlies those desires and intuitions. Within that system, all of our desires point towards happiness. It appears, then, that the evidence on which the experience machine argument relies points towards happiness as the source of prudential value; so the psychological model of happiness and desire conditioning defended above does more than merely allow hedonism to accommodate our experience machine intuitions. Such a model actually suggests that hedonism is true.

This investigation of the process through which our desires and intuitions are formed is also central to my third argument against experience machine intuitions. Those intuitions develop in a world quite unlike the world posited by the thought experiment. They are designed to respond to situations we

face every day, and in those situations they are rather trustworthy guides to our own well-being. They are not reliable, however, when consulted in the fantastical context of the experience machine. This argument does not amount to an outright rejection of the thought experiment. It suggests only that when evaluating the experience machine, we cannot rely on intuitions that were developed in circumstances so far removed from those dictated by the thought experiment.

The experience machine argument fails, therefore, to refute value hedonism. If we are to be persuaded that hedonism is incorrect, we need more than strange thought experiments and an appeal to anti-hedonistic intuitions. According to D.W. Haslett, "we need a reason *why* something that could not possibly make any difference in any of our experiences would nevertheless make us worse off." But, he observes, "a reason is what those who trade in fantastic fictions never give us."⁵⁶ There are indeed other responses to hedonism that purport to give such a reason, although none is as popular or persuasive as the experience machine.⁵⁷ Due principally to Nozick's thought experiment, the past twenty-five years have produced a climate of general hostility towards experientialist accounts of well-being. In this essay I have tried to make the waters a little friendlier. With the experience machine argument no longer an obstacle, hedonism may be poised for a comeback.⁵⁸

Matthew Silverstein

Corpus Christi College, Oxford

matthew.silverstein@corpus-christi.oxford.ac.uk

⁵⁶D.W. Haslett, "What is Utility," *Economics and Philosophy* 6 (1990): 55-94, p. 92.

⁵⁷Three of the more notable arguments against hedonism are the "sour grapes" objection in Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1983); the *Brave New World* argument in Thomas Hurka, *Perfectionism* (New York: Oxford University Press, 1993), p. 27; and the problem of appraisal in Amartya Sen, *Standard of Living*, ed. Geoffrey Hawthorn (Cambridge: Cambridge University Press, 1987), pp. 7-9.

⁵⁸I would like to thank the editors of *Social Theory and Practice* for considering my manuscript, and I am grateful to the anonymous reviewers for their insightful comments and suggestions. Thanks as well to Alexander George, Jonathan Vogel, Derek Parfit, Phyllis Silverstein, Robert Silverstein, Lisa Gourd, Javitt Adili, and especially Joshua Silverstein and Jyl Gentzler. This article would not have been possible without their help.

Copyright of Social Theory & Practice is the property of Florida State University / Dept. of Philosophy and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.