

An Intentionalist Refutation of the “Experience Machine” Debate

CHASE AGHELI

Robert Nozick’s “experience machine” thought experiment focuses on three hypothetical devices that together are able to simulate or effortlessly achieve any feasible human activity: an eponymous “experience machine” that an individual could “plug into” to have any sensation or event they wish to *experience* (Nozick 42); a “transformation machine” that allows one to become anything that they wish to *be*; and a “result machine” that allows one to produce an identical consequence of anything they wish to *do* (44). Nozick argues that if such devices were available to be utilized, the vast majority of people would actually decide *not* to use them, despite hedonistic ethics supposedly dictating that their lives would seem infinitely more pleasurable if they plugged themselves into the machines forever (44). Thus, by demonstrating that most people would be wholly against an activity that has the most hedonistic value out of *any* activity, Nozick argues that hedonism fails to best represent our moral interests.

Each of these machines, in their own way, serves as an argument against a certain kind of effectively limitless bliss, which is often considered the end goal of hedonistic morality. Thus, the thought experiment has

Chase Agheli is a sophomore at Cornell University studying labor relations, philosophy, and cognitive science. His primary philosophical interests include moral psychology, phenomenology, social philosophy, and jurisprudence. After finishing his undergraduate education, he intends to pursue graduate studies in both philosophy and law.

become one of the most infamous arguments against ethical hedonism (Kolber 10). The experience machine is easily the most well-known of these three hypothetical machines and has been the subject of many attempted refutations, but the prevailing trend among such counterarguments is the focus on status quo bias. This counterargument states that we only have familiarity with our current state of affairs (i.e. living in a normal world without instant gratification and where we assume and expect causal relationships for our actions). Thus, we inherently choose our real lives over those offered by the machines only due to the relative normalcy (to us) of the former. However, both Nozick's argument and this counterargument fall victim to the same flaw: a constrained view of pleasure in relation to the non-sensational components of pleasurable experiences. By considering the respective experiences of the machines and real-life as comparably identical, Nozick assumes sensory experience is the only means of creating a positive experience. He neglects the primary distinction between the two—that is, the actual act or sequence of acts involved in creating or realizing a real-life experience. If it is demonstrable that pleasure is also derived from this action or sequence of actions, Nozick's claim of the consistent superiority of the machine can be successfully refuted. In an attempt to refute the thought experiment, the status quo counterargument encounters the same flaw of ignoring how the external pleasures outside of and apart from the machine influence its pleasure output. By doing so, it draws the question: Is the relative comfort that influences an individual to prefer the status quo over the machine not a valid aspect of the status quo's pleasure output? In this paper, I will argue that by framing pleasure within a phenomenological framework of intentionality, we can see that the experience machine actually lacks a source of utility by depriving its user of the real-world directional activity that produces a consequent experience or result. Because the fulfillment of potentiality that separates real-world experience from these machines' simulated reality is in itself a source of pleasure, the hedonistic output of the former is greater, and thus the human preference for true experience is still valid under a hedonistic ethical model. Then, I will examine the chief arguments against the experience machine and show that they, too, neglect the important idea of intentionality in their arguments, such that both sides of this famous argument misinterpret the experience of pleasure in their argumentation.

Nozick's Misinterpretations of Hedonic Pleasure

First, we should determine the discrepancies between Nozick's definition of pleasure, demonstrated in his rationale behind his machines, and the definition generally supported by the hedonist tradition.

One of the foundational flaws of the experience machine thought experiment is its misinterpretation of hedonism intrinsic in its emphasis on unreal experience as a hypothetical maximized source of pleasure. Nozick's logic seems to be that the experience machine is automatically more hedonically "good" than any finite real-world pleasures, because it gives a lifetime of continuous pleasure (albeit of an unreal variety). Mill, in his book *Utilitarianism*, makes an important distinction that shows how Nozick's characterization is false.

To preface, an analysis of the exact kind of pleasure that the experience machine would create is needed. Nozick considers the pleasure exacted by the machine flawed because it is real only in sensation (43). For instance, the experience of being granted an award while one is plugged into the machine is real only in the emotion given; there is no actual work or effort—in Nozick's words, "actual contact" with reality—being done (42). Thus, we can conclude that Nozick is stating that hedonism is flawed because it disregards actual activity and embraces experienced results alone. This argument lines up directly with Mill's example of the archetypal counterargument to hedonism—that sensation encapsulates pleasure because it is the facet of a hedonistic activity that produces a pleasurable sensation, and to live life prioritizing such meaningless sensations is morally problematic (Mill 7). Mill subsequently offers a direct counterargument to this claim. He argues that these base pleasures do not form the entirety of the set of pleasurable experiences, but are part of the set of "lower pleasures," specifically simple, physical pleasures that come about as reactions to experienced stimuli (8). He subsequently posits a corresponding set of "higher pleasures." These, rather than base sensory feelings, require the use of intellect and imagination to produce. They are distinct "in their circumstantial advantages rather than in their intrinsic nature" (8). Through this differentiation, we can see how this superior form of pleasure is unequal to those of the machine: the use of intellect and imagination to *create* pleasure is not present in the machine's experiences, rather they are fully automatic without any necessary effort on the user's part (Nozick 42–43). Furthermore, Mill characterizes higher pleasures in part by their external circumstances, the effects of which can contribute to its hedonic value. The transcendental nature of the experience machine deprives users of these qualities, offering only the base sensations that typify lower pleasures.

Mill's hedonism has no qualms with valuing certain pleasures above others, as long as pleasure as a whole is prioritized and maximized. Mill highlights that people show a greater preference for activities that require real effort and the use of "higher faculties" over meaningless stimulus (9). This leads to an important conclusion regarding the relationship between

the experience machine and real phenomena: if it can be found that an experience offers pleasure of the higher kind, through some action of the higher mental faculties being a factor in its existence, it can be prioritized over lower, solely stimulus-based pleasures, regardless of the quantity of either.

Nozick's Machines as Deprivations of Directionality

Nozick's explanation of the experience machine is matched in significance only by its brevity. An excerpt from Nozick's concise paper highlights his argument: "What is most disturbing about [the machines] is their living of our lives for us... perhaps what we desire is to live (an active verb) ourselves, in contact with reality" (44-45). Here, we see the crux of Nozick's argument: living in pleasure, but without actually performing any tangible or substantial action, is not living actively. To find the origin of this argument, we must examine the foundation of the ethics inherent within Nozick's thought experiment. The ethical doctrines within *Anarchy, State, and Utopia*, the broader treatise in which the machines are postulated, are heavily influenced by Kant, and one segment of Kant's ethics—the teleological argument—is a clear influencing force behind Nozick's arguments with the experience machine. Nozick stresses the distinction of *to live* as an active verb, pointing out that peoples' rejection of the machine means they naturally want "to *do* certain things, and not just have the experience of doing them" (43). Similarly, Kant's teleology stresses purpose over mere happiness (Kant 4:396) and his theory of good will posits the importance of qualification, the assurance that factors such as intelligence be used for good and effective pursuits (4:394). In both maxims, he agrees with (and obviously influenced) Nozick in saying that mental faculties should be directed towards creating real accomplishments and effects; to put it simply, to focus on actual substance over indulging in mere form. And yet, despite Nozick's primarily directional metaethics, his rationale for the experience machine's inferiority compared to real-world activities is purely sensational.

Nozick seems to believe that he includes both a sensationalist (that pleasure is an abstract feeling created solely by sensory input) (*APA Dictionary of Psychology*) and intentionalist (that pleasure is an emotional reaction caused by and directed *towards* something) account in the experience machine experiment (Katz). In fact, as Kolber deduces in "Mental Statism and the Experience Machine," he believes the thought experiment *must* account for all theories of mental states of pleasure, including those that are "non-experiential" (Kolber 11). To do so, Nozick imagines the transformation and result machines. He reasons that, if people shy away

from the experience machine, due to their desire to focus on real outward objects and effects in the world, the ability to do or become whatever they want should fulfill that desire (Nozick 44).

Nozick's counterarguments for the machines refute the superiority of both abstract pleasure and pleasure in instantaneous effect over actual experiences; however, from an intentionalist perspective, these two arguments are incomplete, because both stress the end result of the machines rather than the entire context. Nozick fails to separate the *what* of the simulated experiences or real results from the *how* and *why*. Despite the "emptiness" that forms the center of his argument for why people would prefer reality over the machine, Nozick equates the fully sensationalist (in that the only real aspect is the sensory data one receives) experience of the machine as equally pleasurable as a real-life experience that is directed and fully realized through prior actions and circumstances. This pretense is largely opposed by prominent philosophers of mind like Ryle and Anscombe, "who rejected any account on which pleasure is a context-independent 'internal impression,' whether affective or sensory" (Katz). Ryle and Anscombe's sentiment is similarly found in more scientific approaches to hedonics. For example, in neuroscience, pleasure is in fact considered distinct because base sensory experience is accompanied by "the hedonic marking of affects" (Moccia et al. 2). What is meant by this is that pleasurable sensations cause a consequential and retroactive influence on active behaviors, chief among them *motivation*. For example, if one encounters a food item they have never encountered before, tastes it, and finds that it is pleasurable, they gain new positive memory and motivations (e.g. to find and eat more of that food item). The fact that such experiences cause new motivations and preferences directly related to the item in question (eating more of that specific food) rather than just incentivizing the experiencer to pursue any activity that would grant additional pleasure, shows that pleasure can be directly tied to the affective experience *and pursuit of* a specific object or concept, rather than to the broad experience of pleasure as a whole. All in all, there is a broad consensus in both theoretical and empirical analyses that this activeness Nozick attaches such importance to is a factor of pleasure (and thus, hedonism), not distinct from it. This exposes the foundational flaw in his argument, that he sees doing as separate from feeling. While he sees "doing" as equally important, the experience machine presupposes its value to be separate from pleasure (Nozick 43).

An alternative is offered in Tim Crane's comprehensive intentionalist model of mental states, which goes one step further by arguing that the mind and the object of reaction (in this case the experience, transformation, or result) are not the only elements within the directed relationship proposed

by directed emotion, but the directedness itself is as well: “the conscious character of a state of mind is determined by... its *entire intentional nature*” (Crane 1). He includes within this consciousness factors such as motivation, imagination, and action, all of which echo the importance of intellectual faculties in the pursuit of pleasure espoused by Mill (Crane 14).

Due to these similarities, it is not difficult to reframe Mill’s explanation of hedonism through the lens of intentionality. The higher pleasures, which use intelligence and action to produce an experience beyond basic sensation, are simply the direction of mental faculties into constituting pleasure for an object outside the self. As previously noted, the surrounding context both is *part* of the pleasure experience *and* plays an integral role in constituting it. The act of directional focus to the object outside of the body is just as important as the felt pleasure itself. But, when this outside object does not *actually exist*, and instead is rather just a bundle of sensory variables imitating the object, the act of directionality collapses.

Here is where *activity* comes into play: people do not reject the experience machine because they want something *more than* pleasure, but because they want *more of it*, from a source exclusive to reality. The difference between successfully getting pleasure through an object of intention, and getting pleasure through the imagined end goal of that intention, is the actual act of doing and then *completing* the intended action, an act that reality allows for but the experience machine—by definition—lacks. In this way, the intentionalist model of pleasure explains the rejection of the machine while fulfilling Nozick’s adoption of Kant’s requirement of activity in morally correct actions, as well as both the utilitarian quantification and the emphasis on usage of intellect in Mill’s hedonism.

The previously mentioned secondary thought experiments, the result and transformation machines, are no less flawed in their idea of pleasure as an aspect separate from the human desire to fulfill potentiality. These thought experiments *do* differ from the experience machine: while the experience machine imparts experience without effect, these two do the opposite, creating a desired outcome with no need for any actual action on the part of the machine’s user. We previously argued that individuals’ preference for effect is due to the satisfaction of accomplishing real directionality, thus showing that real world actions possess greater hedonist value than the experience machine. Now, we must explore how this same difference applies to these secondary machines. It is actually fairly simple and can be conceptualized through the exact same structure as the deficit of about-ness possessed by the experience machine. While the experience machine creates *experiences* without effort, these machines create *transformations* and *results* without effort. The directed effort necessary to bring a real pleasure experience into existence is a manifestation of the

directionality that characterizes pleasure intentionalism. Without the directionality inherent in creating the transformation or result, a crucial piece of the pleasure equation is lost, and the products of the machine actually become *less* pleasurable.

Thus, by acknowledging that the act of *achieving* an experience is itself a source of pleasure, we add a new element to the comparison between the output of the machine simulation and that of reality. From here, a proof of its hedonistic superiority over the experiences of Nozick's machine can be created using his own logical rationale found within the same book that the experience machine is introduced. Nozick creates an analytic equation to quantify an individual's decision whether or not to engage in a harmful act. Deterrence is only achieved if expected punitive recourse is more than the expected personal gain from the act, written as " $p \times (C+D+E+R) > (1-p) \times G$," with p being probability, G being positive gain, and all other variables being elements of negative consequence (Nozick 60-61). Ultimately, this equation boils down to a simple case of incentive versus disincentive, but it is important because it shows that Nozick's believes choice based on utility (in this case, personal gain) can be expressed quantitatively. We can thus apply the same method of quantitative modeling to the argument regarding greater pleasure between reality and the experience machine. If it can be shown through this model that reality can have greater hedonic value than the machine, Nozick's claim that our preference for reality contradicts hedonic utilitarianism is refuted.

Nozick's argument can be deconstructed as a very simple inequality $R ? M$, where R is the pleasure derived from reality and M is that derived from the machines (and $?$ is simply a stand-in for the unknown correct inequality operator), which can be solved to determine which experience an individual should choose according to hedonism. Nozick assumes $R < M$ is correct in all cases, because his belief is that real and simulated pleasure experiences are equal in terms of sensation. Thus, the machine always provides greater pleasure by the simple virtue of letting one experience a greater quantity through constant stimulation (if two quantities have equal value per unit, more is always of greater value than less). However, with the knowledge that the directionality of realized action can also contribute to pleasure, the inequality becomes more complicated. We can add pleasure derived from fulfilled intentionality I to the real side, thus creating $(R + I) ? M$. Additional pleasure I can result in the overall pleasure from real experiences equaling or even surpassing M (additional value would presumably differ based on individual preference). While this does not prevent the experience machine from being hedonistically superior in some instances, it does refute Nozick's notion that the experience machine is *always* a more pleasurable experience than reality, thus disproving his

conclusion that the experience machine is a hedonistic moral imperative. The same is true for the secondary machines. Whereas Nozick claims $R < M_T$ and $R < M_R$ for the transformation and result machines, respectively, the pleasure derived from the act of completion of either consequence adds pleasure from completion, C , to the left side of the inequality, rendering Nozick's proposed solution not necessarily true in every circumstance.

In all three cases, reality's hedonic superiority is especially—and arguably consistently—true when Mill's distinction of pleasures in hedonism is taken into account. By virtue of being directional and action-based components of the pleasure experience, C and I add Mill's "higher pleasure" into the inequality. The experience machine remains a lower pleasure due to being strictly a base sensation. As explained previously, Mill's hedonistic utilitarianism asserts that higher pleasures can be prioritized over lower variants. If we return to the example of "value per unit" used above, we can quantitatively represent this as higher pleasure having more value-per-unit than lower pleasures, such that if an experience of higher pleasure and an experience of lower pleasure contain the same "quantitative" amount of pleasure, the former still has greater hedonic value due to having more value *per unit*. Thus, not only does the addition of C or I add more units of pleasure to the real side of the inequality, but these units are more valuable. So, even in cases where the machine possibly grants more pleasurable experience *by unit* (e.g. when comparing the machine's continuous lifetime of pleasure to reality's discrete pleasurable events) the greater value-per-unit of these additions can allow for reality to have more hedonic value overall.

Therefore, we can see through an application of Nozick's own quantitative methodology that the experience machine and its derivations do not fully and completely defeat hedonism as a realistic ethics. The added pleasure of fulfilling potentiality present in real experience shows that the majority's assumed refusal to enter the experience machine is still hedonistically rational.

The Flaws of the Status Quo Bias Argument

In examining the logical flaws within the experience machine, we must also note the oppositional end of the discourse. The counterarguments against the experience machine fall victim to the same misunderstanding of ethical hedonism by accepting Nozick's fundamental premise that the sensationalist experience constitutes the only source of pleasure in an action. While many counterarguments against the experience machine exist—such as that a person plugging themselves in would deprive other people of that person's support, or that the machine might malfunction—they frequently have simple counterarguments that prove the experience machine would still

be the most pleasurable decision. As listed in Bramble's "The Experience Machine," e.g., the experiment assumes that the aforementioned "other people" would have equal access to the machine, or that the machine can be assumed to never malfunction, etc. (139–41). Thus, as these arguments fail to disprove the machine's preferability while still accepting its premise of sensationalism, the majority of these counterpoints can be reliably dismissed. These are refuted within the thought experiment's defined argumentative boundaries without even taking our argument of intentionality into account. However, one remains: the status quo bias argument.

The status quo bias is the most common refutation to the experience machine. It posits that the experience machine is only seen as less preferable because it is disruptive to our baseline perception of activity (Hindriks and Douven 281). Bramble's account defines this as individuals finding "the thought of plugging in too scary, icky, or alien" (139). While Bramble says that this feeling proves Nozick correct in his belief that the experience machine is unpreferable, he fails to address the actual argument. If non-machine experiences were not the status quo, people would not find plugging in "icky."

Despite the wide consensus among most of Nozick's opponents of validity in this counterargument, I instead argue that it does not accurately refute the experience machine, ironically, due to falling into the same fundamental trap as the thought experiment itself. As previously explained, Nozick concludes through his thought experiment that most consider the ability to make actual choices to be an important aspect of life; however, despite what Nozick's thought experiment claims, this does not actually disprove hedonism by showing that people will not choose instant pleasure. Instead, by showing that people desire the complete experiences of real life over the empty experiences of the machine, he inadvertently posits that people—as evidenced by their preference for reality over the simulation—simply believe they will be *more* happy experiencing the former. The fundamental flaw in Nozick's argument is that he establishes his thought experiment on a principle of pleasure versus potentiality, attempting to prove that a rational person choosing to *do* things rather than simply experience them disproves hedonism as a plausible moral code. In truth, it just shows that people find pleasure *in* potentiality and thus would rather complete real experiences. The status quo bias argument follows the same flawed logic in separating two facets of the mental action. In the previous case, this inaccurate separation was between pleasure and the fulfillment of potentiality through the completion of an action, but, in this case, the separation is between pleasure and the decision to continue with the status quo. If the defining objection to the machine under the bias argument is that the act of plugging in and subjecting oneself to false experience

is “alien” and “icky”, then the primary reason to avoid the machine is to avoid an experience of perceived disgust and discomfort. Like many other cases within this debate, this explanation for the bias has empirical support; in a psychological study on the psychology of the experience machine, eighty-seven percent of respondents agreed they would rather stay plugged in then return to reality if they would be subjected to unhappiness and imprisonment in the latter instance (Hindriks and Douven 294, also see 285). In both Bramble’s proposed exception and the empirical study, despite opposing end decisions, the ultimate deciding factor was the option that was less injurious. It can be concluded that one remains in their original experience primarily in order to minimize prospective pain, which—under Mill’s definition—is identical to increasing pleasure, and as such is a primary component of hedonistic utilitarian ethics (Mill 6).

Thus, much like in Nozick’s original argument, the status quo bias argument suffers from a misunderstanding of the precepts of hedonism as pleasure relates to action; however, while the experience machine’s flaw is that it separates the end state of pleasure from the possibility of pleasure through means of attaining the end state, the status quo bias argument separates the hedonistic choice of reality versus the machine from the possibility that the act of living within the status quo, the means taken before the making of the decision, is a source of pleasure by minimizing future pain. In both cases, the action is neglected when, under phenomenological intentionalism, it should be included as part of the hedonistic quantification of pleasure.

Conclusion

In summation, both sides of the debate regarding the experience machine’s defensibility make the mistake of only defining pleasure based on the sensory experience of its end result, and the directedness of the experience, which can be defined as the act of actually creating the experience, is ignored and seen as hedonistically valueless. This rationale is not only contradicted by major findings within experiential philosophy of mind and psychology, but also exposes a key contradiction within Nozick’s own argumentation. His philosophy, which is heavily influenced by Kantian ethics, stresses the importance of “activity” within living a moral life, and yet, he fails to account for the existence of the same valuation of activity within the ethical system he is debating, instead assuming that it is irrelevant to the question of pleasure maximization when it is in fact heavily emphasized by foundational hedonic philosophers. The fact that such a principal argument against hedonism has such flaws opens up far-reaching implications for the value system’s validity and the future of debate in hedonic ethics.

Works Cited

- Bramble, Ben. "The Experience Machine." *Philosophy Compass*, vol. 11, no. 3, 2016, pp. 136–145.
- Crane, Tim. "Intentionalism," edited by Brian McLaughlin & Ansgar Beckermann, *The Oxford Handbook to the Philosophy of Mind*. Oxford: Oxford University Press, 2009, pp. 474–93.
- Hindriks, Frank & Igor Douven. "Nozick's Experience Machine: An Empirical Study." *Philosophical Psychology*, vol. 31, no. 2, 2018, pp. 278–298, doi:10.1080/09515089.2017.1406600.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals*, edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1998. www.search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=55559.
- Katz, Leonard D. "Pleasure." *The Stanford Encyclopedia of Philosophy*, Winter 2016 Edition, edited by Edward N. Zalta, www.plato.stanford.edu/archives/win2016/entries/pleasure/.
- Kolber, Adam Jason. "Mental Statism and the Experience Machine." *Bard Journal of Social Sciences*, vol. 3, Winter 1994/1995, pp. 10–17, San Diego Legal Studies Paper, no. 08-082. SSRN: <https://ssrn.com/abstract=1322059>.
- Mill, J.S. *Utilitarianism*. New York: Barnes & Noble, Inc., 2005.
- Moccia, Lorenzo, et al. "The Experience of Pleasure: A Perspective Between Neuroscience and Psychoanalysis." *Frontiers in Human Neuroscience*, vol. 12, 2018, doi: 10.3389/fnhum.2018.00359.
- Nozick, John. *Anarchy, State, and Utopia*. New York: Basic Books, 2013.
- "Sensationalism." *APA Dictionary of Psychology*, 2007. *American Psychological Association*.