

Nozick's Experience Machine is Dead, Long Live the Experience Machine!

Dr Dan Weijers

danweijers@gmail.com
www.danweijers.com
Philosophy Programme
Victoria University of Wellington

Abstract

Robert Nozick's experience machine thought experiment (Nozick's scenario) is widely used as the basis for a 'knockdown' argument against all internalist mental state theories of well-being. Recently, however, it has been convincingly argued that Nozick's scenario should not be used in this way because it elicits judgements marred by status quo bias and other irrelevant factors. These arguments all include alternate experience machine thought experiments, but these scenarios also elicit judgements marred by status quo bias and other irrelevant factors. In this paper, several experiments are conducted in order to create and test a relatively bias-free experience machine scenario. It is argued that if an experience machine thought experiment is used to evaluate internalist mental state theories of well-being, then this relatively bias-free scenario should be used over any of the existing scenarios. Unlike the existing experience machine scenarios, when this new scenario is used to assess internalist mental state theories of well-being it does not provide strong evidence to refute or endorse them.

1 Introduction

This paper continues an emerging experimental tradition in philosophy; evaluating the responses elicited by thought experiments using both empirical studies and philosophical reflection. The responses evaluated here include those elicited by Robert Nozick's experience machine thought experiment (henceforth Nozick's scenario). Nozick's scenario is widely used as the basis for a 'knockdown' argument against all internalist mental state theories of well-being (Nozick 1974, pp. 42–45).¹ Indeed, most undergraduate philosophy students are introduced to Nozick's experience machine purely for the reason of discrediting these theories.

Perhaps most importantly, Nozick's fantastic scenario has been used extensively to show that the majority of hedonistic theories of well-being must be wrong.² It has been argued, however, that this use of Nozick's experience machine thought experiment is

¹ The following are just a sample of the authors who have stated or implied that the experience machine thought experiment is a knockdown refutation of hedonism or all internalist mental state theories of well-being: Atfield (1987, p. 33), Baggini and Fosl (2007, pp. 74–76), Becker (1992, p. 25), Brink (1989, pp. 223–224), Brülde (2007, pp. 26–29, 33), Bok (2010, pp. 24–28), Darwall (1997, pp. 162, 178), Feldman (2002, p. 615), Finnis (1980, p. 33; 1983, pp. 37–42), Griffin (1986, pp. 9–10), Hausman (2010, p. 329), Haybron (2008, p. 21), Hooker (2000, p. 39), Hurka (2011, pp. 68–70), Jollimore (2004, pp. 333–334), Kagan (1998, pp. 34–36; 2009, p. 253), Kazez (2007, pp. 51–54), Keller (2009, p. 657), Kraut (2007, pp. 124–126), Kymlicka (1990, pp. 13–14), (Nozick 1989, pp. 99–117), Railton (1984, pp. 148–149), Rivera-López (2007, p. 75), Sobel (2002, p. 244), Thomson (1987, p. 41), Tiberius (2004; p. 311, n. 4), Tiberius & Hall (2010, pp. 214–215), van Wyk (1990, p. 109).

² It is often assumed that all hedonistic theories of well-being must be internalist mental state theories, but externalist theories have been suggested. See Fred Feldman's Truth-Adjusted Intrinsic Attitudinal Hedonism (2004, pp. 112–114) for a contemporary example and Weijers (2011a) for discussion of the different types of hedonism. See Weijers (2011b) for an explanation of the Experience Machine Objection to Hedonism.

misguided because it elicits responses that are marred by status quo bias³ and other irrelevant factors (e.g., De Brigard 2010; Kolber 1994; Weijers forthcoming). In light of these arguments, this paper investigates two questions. First, is it possible to create a relatively bias-free experience machine scenario? To which I'll argue that it is. And second, what does this new scenario tell us about the plausibility of internalist mental state theories of well-being? To which I'll argue that it does not provide solid evidence in defense or in support of internalist mental state theories of well-being (which is an improvement on the current state of affairs for internalist mental state theorists).

Hedonistic theories of well-being hold that all pleasure and only pleasure intrinsically contributes positively to well-being (and the opposite for pain). The more plausible versions of these theories define pleasure and pain broadly so that they capture most kinds of enjoyment and suffering (respectively). Internalist mental state theories of well-being hold that only the internal aspects of our beliefs, desires, feelings, and other mental states intrinsically affect our well-being. Internalist mental state theorists about well-being do not dispute that external events, such as winning the lottery, can impact our well-being. However, they would argue that winning the lottery only affects our well-being instrumentally (and only to the extent that it affects the internal aspects of our mental states). For example, according to hedonistic variants of internalist mental state theories about well-being, winning the lottery is usually a good thing, not because winning lots of money is good in and of itself, but because winning lots of money tends

³ Status quo bias is best defined as an inappropriate preference for things to remain the same (Bostrom & Ord 2006).

to make people feel happier.⁴ Any use of the term ‘hedonism’ (or its derivatives) from here on refers to all hedonistic variants of internalist mental state theories of well-being.

First, in Section 2, Nozick’s scenario and the experience machine objection to hedonism are described. Then, in Section 3, the evidence that Nozickian and the other extant experience machine scenarios elicit responses tainted by status quo bias, and other confounding factors, is presented. In Section 4, the results of an experiment on Nozick’s scenario are reported and analysed, revealing good reason to believe that Nozick’s scenario implicitly encourages readers to consider confounding factors. Based on these findings, Sections 5 and 6 discuss ideas for reducing participant’s consideration of these confounding factors and the results of experiments designed to reduce participant’s imaginative resistance, status quo bias, and other confounding factors. Finally, Section 7 discusses the implications of these experimental results for Nozick’s scenario and for internalist mental state theories of well-being.

2 Nozickian Experience Machine Thought Experiments

Nozick originally described the experience machine thought experiment, as follows:

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time, you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life’s experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large

⁴ At least until they become accustomed to their newfound wealth or squander it away and revert back to feeling about as happy as they did before their windfall.

library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your *next* two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. Others can also plug in to have the experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everyone plugs in.) Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* Nor should you refrain because of the few moments of distress between the moment you've decided and the moment you're plugged. What's a few moments of distress compared to a lifetime of bliss (if that's what you choose), and why feel any distress at all if your decision *is* the best one? (Nozick 1974, pp. 42–3, his italics).

The vast majority of people who read Nozick's scenario think that they would choose to remain in reality, including some who previously believed that only how their experiences feel to them affects the quality of their lives. Based on this widespread judgment, the experience machine objection to hedonism seems to show that there is more to the good life for the one living it than how our experiences feel to us on the inside. Furthermore, the small minority of people who claim that they would connect to Nozick's experience machine are often perceived to be not fully engaging with the scenario or just plain wrong about what has value for people. In this way the burden of proof appears to fall squarely on the shoulders of those who would connect to Nozick's experience machine to argue why that would be a prudentially rational thing to do.

A major strength of the experience machine objection to hedonism is how much it concedes to its opponents, while still producing a resounding verdict against them; a life in an experience machine is not described as just slightly more pleasurable than a real

life, but rather as a “lifetime of bliss” in which you can receive any and all of the best experiences possible (Nozick 1974, p. 43). Furthermore, because the lack of direct connection with reality in an experience machine life is assumed to be the only relevant difference between the two options in Nozick’s scenario, most people also infer that living in reality must make our lives go better for us regardless of whether it leads to increased enjoyment or not. It deserves to be emphasised that a key justification for both of these inferences is the widespread dominance of the judgement that connecting to an experience machine in Nozick’s scenario is a bad idea.

Just how widespread is the judgment that connecting to an experience machine in Nozick’s scenario would be worse for us than continuing our normal life in reality? Most introductory ethics lecturers know that the vast majority of students presented with Nozick’s scenario claim not to want to connect to the experience machine. Similarly, most philosophers know that none, or nearly none, of their colleagues would choose to connect to an experience machine.⁵

It is likely that this widespread judgement about Nozick’s scenario has a strong intuitive component. This can be inferred because, upon first exposure, it is not always obvious why the thought of connecting to an experience machine produces a negative

⁵ Initial empirical data from the International Wellbeing Study (IWS) indicates that these judgements may generalise to non-philosophers. In the Further Assessment Study component for the first intake of the IWS, only 12% (19/156) of the participants chose to connect to an experience machine when presented with a simplified version of Nozick’s experience machine thought experiment. The sample for the Further Assessment Study component for the first intake of the IWS (www.wellbeingstudy.com, Jarden *et al.*) is far from representative. It could roughly be described as a self-selecting group of English speakers from several countries around the world who are interested in well-being and like filling out questionnaires about well-being.

feeling.⁶ Unfortunately, intuitive cognition is notoriously opaque.⁷ Since it is impossible to discover what exactly is causing an intuition by introspection, philosophers often attempt to reconstruct the possible causes of their intuitions by reverse engineering the thought experiment in question in order to discover the relevant stimuli it contains. This process of reconstruction often fails because philosophers tend to consider only the factors that are stipulated in the thought experiment as the possible causes of the intuition. But this is a mistake. Intuitive cognition often uses pattern recognition to compare current stimuli with similar past experiences in order to forecast the desirability of the likely outcome (Woodward & Allman 2007). The result of this process is a valenced visceral sensation—usually a good or bad feeling, such as a sinking feeling in your stomach (Woodward & Allman 2007). Since intuitive cognition can use pattern recognition, irrelevant aspects of the triggered past experiences can affect the intuition (Weijers forthcoming). Indeed, even factors that are specifically ruled out by a stipulation in a thought experiment can nevertheless affect our intuitions. This can occur when the irrelevant factors were part of the past experiences that are, on the whole, the most similar to the stipulations of the thought experiment.

For example, a thought experiment that was designed to fairly evaluate a particular kind of life might describe that life as being flawlessly computer-generated to make the

⁶ Since deliberative judgments are open to introspection, and intuitive judgments are not, judgments that appear to be formed *because of* a reason are more likely to have a large deliberative component and judgments that appear to be formed without any immediately obvious reason are more likely to have a large intuitive component. See Weijers (forthcoming) for the difference between deliberative and intuitive cognition in relation to thought experiments.

⁷ Intuitive cognition is opaque because, as a process, it is unconscious until the very end point, when we experience a positive or negative visceral sensation or ‘gut feeling’ (Lieberman 2000; Myers 2002; Weijers forthcoming; Woodward & Allman 2007).

scenario more plausible. The stipulation that the life in question would be flawlessly computer-generated is supposed to be irrelevant, but it is actually quite likely to affect our intuitive judgment of that life. Many of our experiences with complex computerized machinery have involved disappointing underperformance and catastrophic crashing. (Is there an academic who has not lost important work due to her computer crashing?) So, when our intuitive cognition is matching the stipulations of the thought experiment with our past experiences, some of our myriad experiences of computer failure might be the most similar to the stipulations of the thought experiment *as a whole* despite contradicting some individual stipulations. So our intuition about choosing the life that just happens to be flawlessly computer-generated could be influenced by the misapplication of our otherwise rational fear of computer failure.

Regardless of the actual causes of intuitions, when the vast majority of philosophers share an intuition, or when a philosopher holds one so strongly that she assumes it is widespread, that intuition is often used as a premise in philosophical arguments. Most philosophers do not consider widely agreed upon intuitions to be *unquestionable* premises, however. David Sobel, for example, uses the widespread intuition that a real life is better than a life connected to an experience machine as a premise in his argument against quantitative hedonism, while acknowledging that discrediting the intuition would refute his argument (2002, p. 244). Sobel explains that the credibility of intuitions elicited from contemplation of thought experiments can be undermined by “telling a convincing story about the genesis of such intuitions that would explain why we have them while revealing them to be misleading” (2002, p. 244). Apparently unbeknownst to Sobel, a version of this story had already been told by Adam Kolber (1994), who

identified the status quo bias as the main cause of the widespread negative intuition about connecting to an experience machine in Nozick's scenario. Furthermore, this story has since been made more convincing by Felipe De Brigard's (2010) empirical study of reversed experience machine thought experiments and Weijers' (forthcoming) explanation of how our judgements about the experience machine can be easily misled by status quo bias.

3 A Taxonomy of Problems with Experience Machine Scenarios

Since the main goal of this paper is to create a relatively bias-free experience machine scenario, it's worth pausing to take stock of some key ways in which our intuitions about the experience machine might be misleading.

Many philosophers have noted that some of the features of Nozick's scenario might deter readers from connecting to an experience machine because they seem to engender considerable imaginative resistance. People can be said to suffer from imaginative resistance if they consciously, or unconsciously, reject any of the stipulations (or implied features) of a thought experiment.

Of the aspects of Nozick's experience machine thought experiment that have been identified as possible sources of imaginative resistance, nearly all involve how experience machines actually work. Kolber (1994, pp. 13–14), Wayne Sumner (1992, p. 216), and many others have raised concerns that thinking about connecting to an experience machine in Nozick's scenario causes us to worry whether the machine will live up to our expectations or crash and wreak havoc on our lives. It has also been argued that worries about machine failure and underperformance might affect our judgments via our intuitive

cognition, meaning that we may not even be aware of the cause of our imaginative resistance (Weijers forthcoming). Nozick's experience machine sounds like it is a very complicated computerized machine. Based on the ubiquity of experiences of computers crashing, Weijers argued that that most of us will find it very hard to consciously and subconsciously accept that a complex computerized machine could safely look after us (in the manner promised) for the rest of our lives (Weijers forthcoming).

It also seems as though the participants are likely to imaginatively resist Nozick's insistence that we need not worry about our loved ones because they too can plug into an experience machine (Nozick 1974, p. 43). Surely many people would find it hard to disregard their responsibilities to their loved ones, especially given the justification that they could simply be coerced into a major life-changing decision to suit us (Weijers forthcoming).

Still other factors of Nozick's thought experiment might affect our judgment about the value of a life connected to an experience machine because of their tendency to trigger what might be loosely described as the opposite of imaginative resistance—overactive imagination. People can be said to suffer from overactive imagination if they accept the premises of the thought experiment, but make their judgment based on irrationally overblown responses to one or more of those premises. As Kolber suggests, Nozick's description of a tank, electrodes, and having to *plug* in is reminiscent of "science fiction horror stories" (Kolber 1994, p. 14) and likely horrifies people with even mild technophobia.

Status quo bias is the lead villain in Kolber's (1994) story about our intuitive responses to Nozick's scenario, however. Status quo bias is best defined as an

inappropriate preference for things to remain the same (Bostrom & Ord 2006). Status quo bias is closely linked to a family of psychological biases, including loss aversion (valuing losses more than equivalent gains in uncertain circumstances) and the endowment effect (overvaluing what we have and know).⁸ Kolber argues that our intuitions about thought experiments should only be considered good evidence for an argument if other thought experiments considering the same issue elicit similar intuitions (1994, p. 13). Kolber devised a thought experiment that reverses one aspect of Nozick's scenario, while keeping the other aspects consistent and focussing on the same issue (1994, p. 15). Kolber reversed the place from which the decision is made; in his scenario readers discover they are already in an experience machine and are asked if they would like to remain connected or go to reality. As fantastic as this sounds, Kolber reminds us that "[i]t is, in fact, a possibility that you *are* currently hooked up to such a machine, since there is nothing in the world... that could prove otherwise to you" (1994, p. 15, his italics).⁹

Kolber asserts that more people would want to stay connected to an experience machine in his reversed scenario than would agree to connect to one in Nozick's scenario (1994, p. 15). Kolber goes on to argue that the different intuitions elicited by the two scenarios reveal that a bias is likely to be affecting our judgement about what matters to us. Kolber mentions several potential sources of bias, but concludes that continuing the status quo seems to be what really matters to us when we consider experience machine scenarios (1994, pp. 15–16). Status quo bias is implicated in this result because the only

⁸ See for example Druckman (2001), Gilbert (2006), Gilovich, Griffin and Kahneman (2002), Kahneman, Knetsch and Thaler (1991), Kahneman and Tversky (2000), Samuelson and Zeckhauser (1988), Tversky and Kahneman (1991), and Eidelman and Crandall (2012), who provide a useful summary of the psychological research on status quo bias.

⁹ This line of argument is developed further in Weijers (2011c).

difference between the two scenarios seems to be whether reality or the experience machine is framed as the status quo. Since the status quo is arbitrary, Kolber advises against using the experience machine to investigate what should matter to us regarding well-being (1994, p. 16).

Kolber's argument could be easily dismissed if more people would still choose reality over staying in an experience machine in his scenario. However, Filipe De Brigard (2010) provides experimental evidence that reversed experience machine thought experiments elicit very different intuitions from Nozickian ones. De Brigard asked four groups of students whether they would choose an experience machine life or a real life in different scenarios. All of the scenarios were Kolberian reversed experience machine thought experiments, but each described reality differently.

The most relevant of De Brigard's (2010) scenarios is his Negative scenario, which asks respondents to choose between remaining in an experience machine and a real life as a prisoner in a maximum security prison. This is the most relevant of De Brigard's scenarios because it is the closest to Kolber's (1994, p. 15) suggestion and, as Kolber predicted, the result is very different from what we would expect (and I have found) from testing Nozick's scenario; less than 13% (3/24) of the participants responding to the De Brigard's Negative scenario chose reality (De Brigard 2010, p. 47).¹⁰

¹⁰ Unfortunately, De Brigard did not test Nozick's scenario on any of his sample groups, so we can't easily get an idea of exactly how large the difference is between the responses to Nozick's scenario and to his reversals of it. The best we can do is to compare the results of De Brigard's Negative scenario with my test of Nozick's scenario. Less than 13% (3/24) of the participants responding to the De Brigard's Negative scenario and about 84% (66/79) of participants responding to my test of Nozick's scenario reported preferring to live in reality. This difference is certainly large, but it's impossible to say how much of the difference is caused by varying characteristics between the sample groups and other experimental conditions.

De Brigard's Negative scenario also appears to be the most similar to Nozick's scenario in terms of the relative difference between the two key aspects that experience machine thought experiments are widely thought to compare (reality and how our experiences feel to us on the inside).¹¹ In Nozick's scenario, the choice is between average¹² but real experiences and great but unreal experiences. In De Brigard's Negative scenario, the choice is between bad but real experiences and average but unreal experiences. Even if we compare my Self scenario (which is very similar to Nozick's scenario, except that it appeared to reduce the impact of irrelevant factors)¹³ with De Brigard's Negative scenario, there is still a very large difference between the respondents' reported choices. De Brigard believes that whether reality or an experience machine was framed as the status quo explains a considerable amount of this difference between the responses to his scenarios and the responses to (non-reversed) Nozickian experience machine scenarios (2010, pp. 50–51).

De Brigard's experiments have been criticized,¹⁴ but these criticisms should not be taken as complete refutations of his conclusions, especially when the wealth of support

¹¹ Where 'reality' refers to living in direct contact with reality in a way that we can freely interact with it.

¹² Nozick doesn't explicitly state that your life has average experiences, but I'm assuming that most people who read Nozick's scenario are roughly average in experiences.

¹³ My Self scenario is discussed in more detail in Section 5.

¹⁴ Smith (2011) criticizes De Brigard's (2010) experiments on several grounds, including: the sample groups being too small, a lack of information about the procedure of the experiment, and a significant disanalogy between De Brigard's scenarios and Nozick's scenario. The disanalogy, which involves an imbalance in how relationships with others and social capital is treated, is also likely to affect judgments about Kolber's (1994) scenario. Smith's own experiments demonstrate that the disanalogy may significantly enhance the appearance of status quo bias in De Brigard's results. But even if this affect could be taken into account, De Brigard's results would still indicate the respondents had been non-negligibly influenced by status quo bias. Smith also expresses concern with the representativeness of De Brigard's all-student sample, but since this is also a potential issue for my experiments, it is addressed in Note 37 below.

for the ubiquity of status quo bias in the social sciences literature is taken into account. The considerable number of studies establishing the broad influence of status quo bias might be resisted if they were all experiments concerning swapping chocolate bars and mugs,¹⁵ or even experiments concerning important financial decisions.¹⁶ The fact of the matter is, however, that even *post facto* analysis studies of real life important decisions (such as choosing a mutual fund) also demonstrate the pervasive effects of status quo bias.¹⁷ Even establishing torture as the status quo (by saying that it has been common practice for 40 years) made participants in one study significantly more likely to support the practice of torture (Crandall *et al.* 2009).¹⁸ De Brigard's (2010) and Smith's (2011) results could be explained by several biases or other factors, including chance, but the established pervasiveness of status quo bias offers considerable support for the claim that the effect of status quo bias on our responses to experience machine thought experiments considerably reduces their relevance to questions about well-being.

While it's not clear exactly how much influence status quo bias and other confounding factors have on our judgments about Nozick's scenario, it seems very likely that they are having at least some effect. Therefore, there is a strong case that the normative significance of the intuitions elicited by Nozickian experience machine

Finally, Smith (2011) also argues that all surveys on experience machine thought experiments are destined to fail because survey respondents cannot fully adopt the mental states of an agent confronted with such an important decision. Weijers (2012) argues against this objection at length, so it will not be discussed here.

¹⁵ (e.g., Knetsch 1989; Knetsch & Sinden 1984).

¹⁶ (e.g., Hartman, Doane & Woo 1991).

¹⁷ (e.g., Kempf, A., & Ruenzi, S. 2006; Samuelson and Zeckhauser 1988).

¹⁸ It should be noted that this result might be partially explained by some sort of epistemic laziness based on believing what others believe instead of being completely based on the status quo bias. Thanks to an anonymous reviewer for pointing this out.

thought experiments are undermined. Furthermore, since reversed experience machine thought experiments such as De Brigard's (2010) also frame certain aspects of one of the choices as the status quo, the normative significance of the intuitions elicited by these kinds of experience machine thought experiment is also undermined. Indeed, because familiarity and other irrelevant factors seem to have large effects on our decisions about what life to choose in experience machine thought experiments, these choices should probably be considered too biased to be used as evidence in arguments about well-being.

4 Analysing Nozick's Scenario

To investigate these stories about the likely influences on judgments about Nozick's experience machine thought experiment, I conducted a paper-based survey experiment on first year students at Victoria University of Wellington.¹⁹ The survey consisted of the

¹⁹ The experiment consisted in handing out surveys to four different classes of first year students on 15 August 2012. The surveys were handed out at the beginning of two lectures and the end of the other two. Each survey sheet contained one of seven different experience machine scenarios. No participants responded to more than one scenario and participants who had completed earlier versions of one of the scenarios in a previous experiment were not counted. I was unknown to the vast majority of the respondents and did not communicate with them before or during the experiment except to tell them that participation was not mandatory, not linked to their grades, and completely anonymous, etc. Two classes were marketing and two were philosophy. Using a 1-tailed Fisher's Exact test, I found no statistically significant differences between the results for philosophy students and marketing students (the difference that was closest to being significant was for responses to Nozick's scenario: p -value = 0.143). This methodology assumes that we expect to see a relationship (a difference in the particular direction that was seen in the results) and this result means that we can only be about 85% confident that the difference between the groups is not a product of chance. A 95% level of confidence is usually required to consider the result "statistically significant". Furthermore, since the differences between the philosophy and marketing groups vary in different directions for different (but relatively similar) scenarios and we lack a theory of why we would expect there to be a difference between these groups *in a particular direction*, a 2-tailed Fisher's Exact test should probably be used. Using a 2-tailed Fisher's Exact test, we observe: p -value

exact quote of Nozick's original experience machine scenario from above and the following questions:²⁰

1) What is the best thing for you to do *for yourself* in this situation?

Tick only one of these options:

- Permanently plug in to an Experience Machine
- Never plug in to an Experience Machine

2) Briefly explain your choice:

79 survey sheets on Nozick's scenario were completed. About 16% (13/79) of the students indicated that they would connect to an experience machine. For the 84% of students who didn't want to connect, 'reality, truth, real autonomy, or something related' was the most common justification. However, this justification only accounted for 47% (27/66) of the 'informational main reasons' (henceforth 'main reasons' or 'main justifications') this group gave for choosing reality over the experience machine. 'Informational reasons' are those not coded as un-informational reasons (i.e.: either 'no answer' or 'unclear, incoherent, or un-informational'). Responses coded as 'unclear, incoherent, or un-informational' were generally long-winded versions of "I don't know why". 'Main reasons' refers to the only reason provided, or the first coherent reason if multiple reasons were provided.²¹

= 0.229, which means that we can only be about 77% confident that the difference between the groups is not a product of chance.

²⁰ There was also a third question, asking their gender, for a separate research project.

²¹ 12% (8/66) of the survey forms on Nozick's scenario provided un-informational main justifications.

Surprisingly, 34% (20/66) of the main justifications given by the students who didn't want to connect indicated that they experienced imaginative resistance to the stipulated or implied features of the scenario. The main justifications displaying imaginative resistance were grouped as follows: 'bad experiences are required to appreciate good experiences or to develop properly' (17%, 10/66; 'unpredictable or surprising experiences are better than pre-programmed ones'²² (10%, 6/66); 'I can't because I have responsibilities to others' (3%, 2/66); and 'the machine might break down or not produce great experiences in the future' (3%, 2/66).

Some main justifications also revealed overactive imagination, such as 'the machine seems scary or unnatural' (10%, 6/66). Other main justifications were consistent with the thought experiment, and didn't indicate overactive imagination, but were nonetheless irrelevant for evaluating the relative intrinsic prudential value of reality and how our experiences feel to us on the inside. These main justifications reveal problems with using Nozick's scenario (as it often is) to investigate the relative intrinsic prudential value of reality and how our experiences feel to us on the inside. Most prominent of these confounding factors was: 'getting out every two years would be depressing' (9%, 5/66). Taken together, these results show that many of the factors that have concerned philosophers might well be influencing at least some participants' judgments about Nozick's scenario.

Considering that Nozick's stated purpose for his experience machine thought experiment was to assess what matters, if anything, other than how our experiences feel to us on the inside, all of the justifications for not connecting (from the previous two

²² These main justifications demonstrated imaginative resistance because their experiences would seem unpredictable and surprising while connected to an experience machine.

paragraphs) are irrelevant. The fact that 47% (31/66) of the participants that chose reality (and provided an informational main justification) stated an irrelevant reason as the main justification for their choice gives us reason to believe that Nozick's scenario might not be very useful for assessing the relative intrinsic prudential value of reality and how our experiences feel to us on the inside.²³

5 Reducing Imaginative Resistance and Other Irrelevant Factors

Since all of the extant experience machine scenarios appear to engender so many biased, irrelevant, and imaginatively resistant judgments, a question arises: Is it possible to create an experience machine scenario that is not so heavily affected by these confounding factors? To begin investigating this question, I created a new experience machine scenario that addressed some of the concerns just raised with Nozick's scenario. This new scenario was then tested on students using the following survey:²⁴

It's 2062 and you are riding a hovertube to town. You have been offered a permanent spot in an Experience Machine and are now trying to decide if you should accept.

You have had a go in an Experience Machine before and know that they provide an unpredictable roller-coaster ride of remarkable experiences. When in the machine, it still felt like you made autonomous decisions and occasionally faced tough situations, such as striving for your goals and feeling grief, although you didn't *really* do these things. Your experiences

²³ These stated justifications might not be the real reason for the choice the participants actually made because the main cause of their judgment could have been subconscious—see Weijers (forthcoming) for more on this. Regardless, the justifications given match up to the reasons that some philosophers have predicted would impact participants' judgements and it might be the case that the judgments of the participants who confabulated their justification are somewhat offset by those of the participants who couldn't articulate the real cause of their judgement.

²⁴ See Note 19 for more details about the experiment.

were also vastly more enjoyable and varied in the machine. You also recall that, while you were in the Experience Machine, you had no idea that you had gotten into a machine or that your experiences were generated by a machine.

If you accept the spot, then you would stay in an Experience Machine permanently. If you reject the spot, then you would never be offered a spot again. Your life would be the same length in an Experience Machine as it would otherwise have been.

1) Ignoring how your family, friends, any other dependents, and society in general might be affected, and assuming that Experience Machines always work perfectly, what is the best thing for you to do *for yourself* in this situation?

Tick only one of these options:

- You should accept the spot in an Experience Machine
- You should *not* accept the spot in an Experience Machine

2) Briefly explain your choice:

This scenario will be referred to as the Self scenario. 79 survey sheets on the Self scenario were completed. About 34% (27/69) of participants responding to the Self scenario chose to connect to the experience machine. Comparing the responses to the Self scenario with the responses to Nozick's scenario reveals a statistically significant 18% increase in participants choosing to connect to the experience machine in the Self

scenario.²⁵ Furthermore, since participants were asked to explain their choice, the total amount of irrelevant main justifications can also be compared. In the Self scenario, 31% (16/44) provided an irrelevant main justification for their choice, compared to 47% (31/66) in Nozick's scenario. Most of this difference is explained by far fewer respondents to the Self scenario justifying their choice with the 'bad experiences are required to appreciate good experiences or to develop properly', 'unpredictable or surprising experiences are better than pre-programmed ones', and 'getting out every two years would be depressing' reasons. Although we cannot expect the stated main justification for the participants' choices to correlate perfectly with the actual main reason for their choice, these results nevertheless give us reason to think that the Self scenario is probably a better test than Nozick's scenario of the relative intrinsic prudential value of reality and how our experiences feel to us on the inside.

What could we conclude about hedonism and other internalist mental state theories of well-being based on the results of the Self scenario? It's not clear that we can draw any firm conclusions. Nearly two thirds of respondents to the Self scenario still chose to forego the experience machine. However, all of the 20% (16/79) of respondents who provided an irrelevant main justification for their choice chose reality over to an experience machine life. It would be reasonable to expect that at least some of these participants might have reported preferring to connect to the experience machine if they had not considered the irrelevant factor they cited as justifying their choice.²⁶ Recall that

²⁵ Using a 1-tailed Fisher's exact test, we observe (p -value = 0.008), which means that we can be more than 99% confident that the difference in the responses between these two scenarios is not the product of chance (when we assume that there will be a difference in the particular direction it was in the results).

²⁶ This assumption will not sound reasonable to anyone who thinks that the vast majority of these justifications are likely to be confabulations—vain attempts by our deliberative mind to explain the verdict

one of the main strengths of the experience machine objection to hedonism is that the overwhelming majority agree that it's best to choose reality over an experience machine life. The results of my surveys on Nozick's and the Self scenarios show that this strength might be lost if the participants adhered to the stipulations of the thought experiments more closely. And this is before any attempts to reduce the likely effects of status quo bias.

6 Reducing Status Quo Bias

Assuming status quo bias is interfering with the usefulness of our intuitive responses to most kinds of experience machine thought experiments, the question arises: Is it possible to create an experience machine scenario that is relatively unlikely to elicit responses affected by status quo bias? It should be noted that any successful attempt to reduce the affects of status quo bias on responses to experience machine scenarios provides additional evidence that they were tainted by status quo bias in the first place.

How could a thought experiment be designed in order to reduce the potential effects of status quo bias? One potential way to mitigate the effects of status quo bias is to reduce any oversensitivity to potential losses. This is relevant here because such an oversensitivity would encourage us to prefer our current life to any life that we are not entirely familiar with, such as an experience machine life. Evolutionary considerations give us good reason to think that we should be loss averse in conditions of uncertainty because losses were probably more deleterious to our ancient ancestors' evolutionary

of our intuitive cognition. Readers with this belief needn't accept this part of the argument because they are likely to be more heavily swayed by the main argument about the affects of status quo and other biases on our judgments about experience machine scenarios.

fitness than equivalent gains were beneficial (Chen, Lakshminaryanan, & Santos 2006). Based on this evolutionary consideration, it is reasonable to expect that the less we know and care about someone, the less likely we are to be oversensitive to the risk of them losing something. Indeed, this has been demonstrated in several experiments.

The experiments show that when we make decisions for people in conditions of uncertainty our value function for gains and losses becomes increasingly flat the less we care about the people we are making decisions for (Bloomfield *et al.* 2006). That is to say that we are more likely to value equivalent losses and gains equally, just like an economist's rational agent would, if we are impartial towards the people we are making decisions for. Therefore, I hypothesise that participants will experience less of an inappropriate preference to maintain the status quo, and less overactive imagination and imaginative resistance to stipulations about experience machines, if they are judging the value of the potential lives of someone other than themselves. If this hypothesis is correct, then we should expect that we are more likely to think that the unfamiliar (and therefore risky) experience machine life is going to be a better idea for someone we care less about than ourselves; that participants will be more likely to think that a stranger should connect to an experience machine than a cousin or friend, and a cousin or friend more than themselves.

This hypothesis was tested by adapting the Self scenario to create versions about people that participants should have a weaker emotional attachment to than themselves. These new scenarios involved making the decision about a friend, a cousin, and a stranger.²⁷ The Stranger scenario read as follows:

²⁷ See Note 19 for more details about the experiment.

It's 2062 and you are riding a hovertube to town. A stranger sits down next to you, introduces himself as Boris, and tells you that he has been offered a permanent spot in an Experience Machine. Although you would never actually tell Boris your opinion, you are trying to decide if you think he should accept.

You have had a go in an Experience Machine before and know that they provide an unpredictable roller-coaster ride of remarkable experiences. When in the machine, it still felt like you made autonomous decisions and occasionally faced tough situations, such as striving for your goals and feeling grief, although you didn't *really* do these things. Your experiences were also vastly more enjoyable and varied in the machine. You also recall that, while you were in the Experience Machine, you had no idea that you had gotten into a machine or that your experiences were generated by a machine.

If Boris accepts the spot, then he would stay in an Experience Machine permanently. If he rejects the spot, then he would never be offered a spot again. Boris' life would be the same length in an Experience Machine as it would otherwise have been.

1) Ignoring how Boris' family, friends, any other dependents, and society in general might be affected, and assuming that Experience Machines always work perfectly, what is the best thing for Boris to do *for himself* in this situation?

Tick only one of these options:

- Boris should accept the spot in an Experience Machine
- Boris should *not* accept the spot in an Experience Machine

2) Briefly explain your choice:

The Stranger scenario is very similar to the Self scenario, the only differences being that the choice is now about whether Boris, the stranger, should connect to an experience machine, instead of participants choosing for themselves. 82 survey sheets on the Stranger scenario were completed. As predicted, participants responding to the Stranger scenario were much more likely to think that connecting to the experience machine was a good idea than those responding to the self scenario. Over 52% (43/82) of participants responding to the Stranger scenario decided that connecting to an experience machine made the life in question better, 18% more than participants responding to the Self scenario.²⁸

The results for the Friend and Cousin scenarios add further weight to the hypothesis that status quo bias can be reduced by making the scenario about someone the decision-maker is less emotionally attached to. The Friend scenario makes the subject the participants' (unnamed) friend, instead of a stranger (named Boris), and the cousin scenario makes Boris the participants' cousin (named Boris).²⁹ Otherwise the scenarios remain the same as the Self and Stranger scenarios. It is reasonable to expect that most participants will be more sensitive to the risk of losses to these people in order of how

²⁸ Using a 1-tailed Fisher's exact test, we observe (p -value = 0.015), which means that we can be nearly 99% confident that the difference in the responses between these two scenarios is not the product of chance (when we assume that there will be a difference in the particular direction it was in the results).

²⁹ The exact changes in the Cousin scenario (from the Stranger scenario) are as follows: The second sentence in the scenario is replaced with: "You have just heard that your cousin Boris has been offered a permanent spot in an Experience Machine." The exact changes in the Friend scenario (from the Stranger scenario) are as follows: The second and third sentences in the scenario are replaced with: "You have just heard that your friend has been offered a permanent spot in an Experience Machine. Although you would never actually tell him your opinion, you are trying to decide if you think he should accept." All of the remaining six mentions of "Boris" in the Stranger scenario are replaced with "Y/your friend" in the Friend scenario.

close they are to them, i.e.: themselves, then their unnamed friend and their cousin Boris, and finally a stranger named Boris. Figure 1 (below) shows the differences in the responses to these four scenarios and Table 1 (below) shows the differences between these scenarios and the statistical significance of those differences (calculated using a 1-tailed Fischer’s exact test).

Figure 1: Reducing Status Quo Bias in Responses to Experience Machine Thought Experiments by Decreasing the Emotional Attachment to the Subject

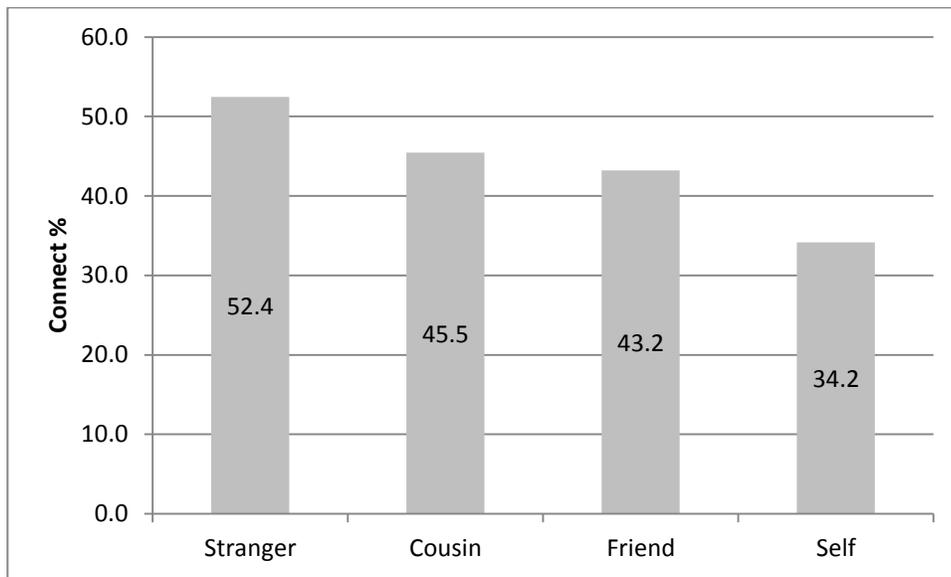


Table 1: Relative Differences in the Propensity to Report that Connecting to an Experience Machine is Better for Well-being than Living in Reality in Scenarios with Differences in the Decision-Makers’ Emotional Attachment to the Subject

Scenarios	Difference in Connect %	Connect #/n & Connect #/n	p-value (3 d.p.)
Stranger & Self	18.3	43/82 & 27/79	0.015
Cousin & Self	11.3	35/77 & 27/79	0.101
Stranger & Friend	9.2	43/82 & 35/81	0.153
Friend & Self	9.0	35/81 & 27/79	0.156
Stranger & Cousin	7.0	43/82 & 35/77	0.235
Cousin & Friend	2.2	35/77 & 35/81	0.451

Table 1 shows that we can only be very confident (about 99% confident) that the difference between the responses to the Stranger and the Self scenarios are not the product of chance. Differences that were not statistically significant by regular standards (p -value < 0.05), but were close, include the difference between the responses to the Cousin and the Self scenarios (which we can be about 90% confident was not the product of chance) and the Friend and Self scenarios (which we can be about 85% confident was not the product of chance). Taken together, these results show that we have good reason to believe that something about the thought of a stranger connecting to an experience machine is more palatable than the thought of connecting ourselves to one. They also show that larger samples would be needed to be very confident that moderately decreasing decision-makers' emotional attachment to the subject of experience machine thought experiments makes connecting to an experience machine seem more palatable.

It might be argued that it is good to be sensitive to risk, especially in important decisions like how to spend the rest of your life. The point of this argument is that eliminating sensitivity to risk is not a virtue of my scenarios, but a weakness. On the contrary, experience machine scenarios should be designed to eliminate *oversensitivity* to the risk of losses because the losses and gains should be fixed so as to better isolate the

relative intrinsic prudential value of reality and how our experiences feel to us on the inside. Furthermore, many of the specific kinds of losses envisaged by participants responding to Nozick's scenario (machine failure, the experience machine's inability to provide all-important negative experiences, etc.) are either specifically ruled out in the scenario or are clearly irrelevant to comparing reality with how our experiences feel to us on the inside.

Another potential method for reducing status quo bias is to frame all of the options as equally familiar. Nozick's scenario, and all of my scenarios so far, are likely to suffer from the status quo bias because they all frame reality as the status quo, and therefore, as a much less risky option. Furthermore, the endowment effect of the status quo bias (overvaluing what we have and what we know) could be reduced by neutralising the status quo. Therefore, I hypothesise that framing both options as equally familiar will produce results that are much less affected by status quo bias. This hypothesis was tested by creating a version of the Stranger scenario in which neither an experience machine life or a real life was framed as the status quo. The Stranger No Status Quo (Stranger NSQ) scenario read as follows:

A stranger, named Boris, has just found out that he has been regularly switched between a real life and a life of machine-generated experiences (without ever being aware of the switches); 50% of his life has been spent in an Experience Machine and 50% in reality. Nearly all of Boris' most enjoyable experiences occurred while he was in an Experience Machine and nearly all of his least enjoyable experiences occurred while he was in reality. Boris now has to decide between living the rest of his life in an Experience Machine or in reality (no more switching).

You have had a go in an Experience Machine before and know that they provide an unpredictable roller-coaster ride of remarkable experiences. When in the machine, it still felt like you made autonomous decisions and occasionally faced tough situations, such as striving for your goals and feeling grief, although you didn't *really* do these things. Your experiences were also vastly more enjoyable and varied in the machine. You also recall that, while you were in the Experience Machine, you had no idea that you had gotten into a machine or that your experiences were generated by a machine.

Boris' life will be the same length in an Experience Machine as it would in reality. No matter which option Boris chooses, you can be sure of two things. First, Boris' life will be very different from *your* current life. And second, Boris will have no memory of this choice and he will think that he is in reality.

1) Ignoring how Boris' family, friends, any other dependents, and society in general might be affected, and assuming that Experience Machines always work perfectly, what is the best thing for Boris to do *for himself* in this situation?

Tick only one of these options:

- Boris should choose the Experience Machine life
- Boris should choose the real life

2) Briefly explain your choice:

As expected, when neither reality nor the experience machine were framed as the status quo, participants were more likely to think that Boris should connect to an experience machine than when reality was framed as the status quo. But the effect was small and not statistically significant. About 55% (42/77) of participants responding to the Stranger

NSQ scenario thought Boris should connect to an experience machine, which is 2.1% more than in the Stranger scenario.³⁰ However, in a previous version of this experiment,³¹ there was a much more statistically significant 12% difference between the Stranger NSQ' and Stranger' scenarios.³² Since the only difference between the Stranger NSQ' and Stranger' scenarios was the framing of the status quo, there might be good reason to think that the main cause of this difference was the reduction of the endowment aspect of the status quo bias—the overvaluing of what we have and what we know. However, this result was not replicated by the difference between the results for the Stranger NSQ and Stranger scenarios of the current experiment. So, it seems that more research is required to be confident that the attempt to neutralise the status quo in the Stranger NSQ scenario was successful.

³⁰ Using a 1-tailed Fisher's exact test, we observe (p -value = 0.457), which means that we can be nearly 55% confident that the difference in the responses between these two scenarios is not the product of chance (when we assume that there will be a difference in the particular direction it was in the results).

³¹ The previous version of this experiment was conducted in 2011 on first year philosophy and business students at Victoria University of Wellington. The only difference in the relevant scenarios was the treatment of autonomy. The Stranger NSQ' and the Stranger' scenarios, included the following text: "When in the machine, you still made autonomous decisions and faced tough situations, such as striving for your goals and feeling grief, but your experiences were vastly more enjoyable and varied." This text was replaced in all of my scenarios discussed in this article with the following: "When in the machine, it still felt like you made autonomous decisions and occasionally faced tough situations, such as striving for your goals and feeling grief, although you didn't *really* do these things. Your experiences were also vastly more enjoyable and varied in the machine."

³² Over 60% (75/124) of participants responding to the Stranger NSQ' scenario thought Boris should connect to an experience machine, compared to about 48% (45/93) of participants responding to the Stranger' scenario. Using a 1-tailed Fisher's exact test, we observe (p -value = 0.051), which means that we can be about 95% confident that the difference in the responses between the Stranger NSQ' and the Stranger' scenarios is not the product of chance (when we assume that there will be a difference in the particular direction it was in the results).

Nevertheless, by comparing the Self scenario with the Stranger NSQ scenario, we can see a 20% increase in the proportion of participants indicating that they think an experience machine life is better than a real life.³³ Assuming that the main reason for this dramatic increase is due to a reduction in the effects of status quo bias, then we have good reason to believe that scenarios that frame reality as the status quo and have the reader, instead of a stranger, as the subject are likely to heavily bias the results in favour of reality. However, this result does not go so far as to clearly support De Brigard's (2010) claim that status quo bias renders Nozick's scenario relatively useless. After all, even a 20% difference in the base population's responses would still mean that the clear majority of participants would not endorse connecting to an experience machine. But when the likely effects of the status quo bias are combined with the other likely confounding factors discussed in Section 5, Nozick's scenario is starting to look a lot less useful than it is commonly thought to be.

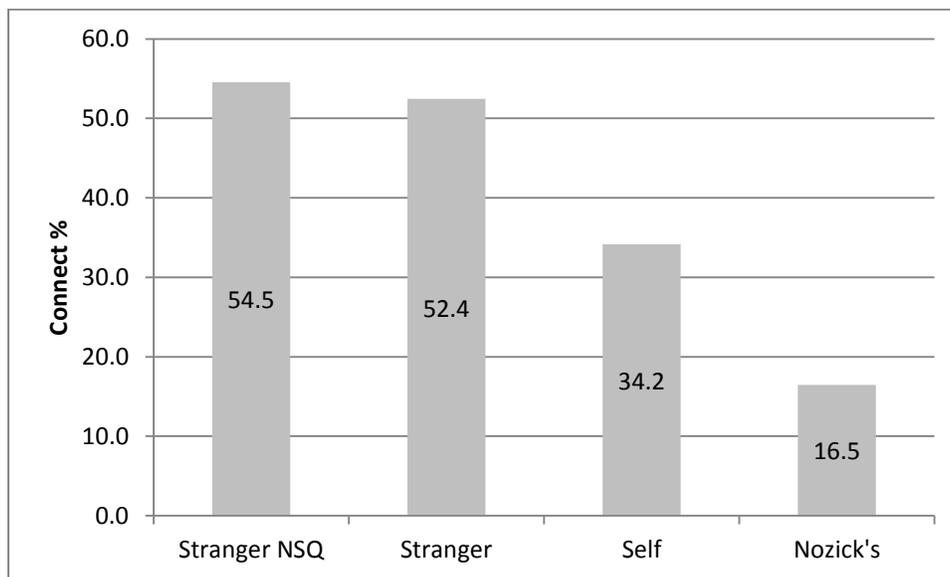
7 Nozick's Experience Machine is Dead, Long Live the Experience Machine!

Assuming that the three main hypotheses put forward so far are well supported by the evidence from my experiments, we can explain the three main reasons for the difference between the high connection rate for the Stranger NSQ scenario and the low rate for Nozick's scenario.³⁴ Figure 2 (below) shows the differences in the responses between my main scenarios.

³³ Using a 1-tailed Fisher's exact test, we observe (p -value = 0.008), which means that we can be over 99% confident that the difference in the responses between these two scenarios is not the product of chance (when we assume that there will be a difference in the particular direction it was in the results).

³⁴ I say 'assuming' here because it is quite possible that I have inadvertently elicited some additional biases or other irrelevant factors in my new scenarios that were not existent in Nozick's scenario. Furthermore, it

**Figure 2: Reducing Status Quo Bias and Other Confounding Factors
in Experience Machine Thought Experiments**



First, we can see that reducing the impact of irrelevant factors, such as worries about needing bad experiences that wouldn't be available from an experience machine and what it would be like to disconnect from the machine every two years, makes about 18% difference (compare Self to Nozick's). Second, we can see that making the choice on behalf of a stranger instead of ourselves also makes about 18% difference (compare Stranger with Self). As was argued, this is presumably because we are less irrationally

might be argued that my all-student samples are unrepresentative of reasonable people generally. However, Smith's (2011) test of his Pretend Neutral scenario on a two-country, multi-setting, and multi-demographic group of participants produced a fairly similar result to De Brigard's (2010) test of his analogous Neutral Status Quo Emphasised scenario on a mono-country, mono-setting, and all-student group (71% and 59% respectively). Furthermore, most of the 12% difference between these two results might be explained by the extra emphasis that Smith put on having to start life anew in reality in his Pretend Neutral scenario. Therefore, there is at least some evidence to believe that student samples are fairly representative of reasonable people generally and that my results cannot be dismissed out of hand for being unrepresentative.

loss averse (oversensitivity to the risk of loss in conditions of uncertainty) when deciding on behalf of people we are less emotionally attached to. Third, we can see that neutralising what was framed as the status quo (the inappropriate bias in favour of what we are familiar with) might be responsible for the 2% difference between the responses to the Stranger NSQ and Stranger scenarios. Furthermore, since a previous experiment also yielded a 12% difference by attempting to neutralise the status quo in the same way, additional research might provide stronger reason to think that neutralising what was framed as the status quo could make a significant difference to responses to experience machine scenarios. Finally, the total difference between the Stranger NSQ scenario and Nozick’s scenario is 38%, a large and very highly statistically significant difference. The differences between the scenarios and the statistical significance of those differences (calculated using a 1-tailed Fischer’s exact test) are displayed in Table 2 below.

Table 2: Relative Differences in the Propensity to Report that Connecting to an Experience Machine is Better for Well-being than Living in Reality in the Main Scenarios

Scenarios	Difference in Connect %	Connect #/n & Connect #/n	p-value (3 d.p.)
Stranger NSQ & Nozick's	38.1	42/77 & 13/79	0.000
Stranger & Nozick's	36.0	43/82 & 13/79	0.000
Stranger NSQ & Self	20.4	42/77 & 27/79	0.008
Self & Nozick's	17.7	27/79 & 13/79	0.008
Stranger NSQ & Stranger	2.1	42/77 & 43/82	0.457
Stranger & Self	18.3	43/82 & 27/79	0.015

The combination of the large differences between the responses to my scenarios and Nozick's scenario with the higher level of imaginative resistance to Nozick's scenario provides good reason to think that Nozick's scenario is not the best scenario to choose when trying to argue about the relative intrinsic prudential value of reality and how our experiences feel to us on the inside. So it seems that Kolber (1994), De Brigard (2010), and Weijers (forthcoming) were correct in calling for an end to the use of Nozick's scenario for evaluating internalist mental state theories of well-being. Indeed, we now have good reason to think that Nozick's scenario elicits intuitions that are about 38% off the mark, and since alternative and seemingly less biased scenarios are available, the use of Nozick's scenario as a knockdown argument against hedonism or any other internalist mental state theories of well-being should be well and truly over. Nozick's experience machine thought experiment is dead!

Assuming that experience machine thought experiments provide useful tools for helping us to address questions about well-being,³⁵ it makes sense to ask which experience machine scenario is best for evaluating internalist mental state theories of well-being. The Stranger NSQ scenario addresses the question of the relative intrinsic prudential value between reality and how our experiences feel to us on the inside while

³⁵ An alternate option to experience machine thought experiments for addressing questions about well-being might be to ask simple questions like 'if you never directly or indirectly experience the effect of an event, can that event make your life go better or worse for you?' In my experience of asking these kinds of questions to anyone other than trained philosophers, they confidently offer answers that often contradict answers that they would give to related questions, such as 'if you are insulted behind your back, but never directly or indirectly experience any of the effects of the insult, does being insulted in this way make your life go worse for you?' Since an example provides more information than just a principle, many philosophers believe that thought experiments can elicit more considered responses than consideration of principles in isolation. Thank you to an anonymous reviewer for bringing this to my attention.

making use of the memorable, mysterious and intriguing concept of the experience machine. Furthermore, the Stranger NSQ scenario appears to be much less affected by the status quo bias and other irrelevant factors without being much more complicated than Nozick's version.³⁶ All of the extant experience machine scenarios, except for the Stranger NSQ scenario appear to be non-negligibly affected by status quo bias and other irrelevant factors. Therefore, although the Stranger NSQ scenario is possibly affected by some problems, we have good reason to believe that it is relatively bias free compared to all of the other scenarios. For these reasons, the Stranger NSQ scenario should be used, instead of any of the extant experience machine scenarios, for investigating the relative intrinsic prudential value of reality and how our experiences feel to us on the inside. Long live the experience machine!

So what could the results of the Stranger NSQ scenario tell us about the relative intrinsic prudential value of reality and how our experiences feel to us on the inside? And what should we think about hedonism and other internalist mental state theories of well-being? Since about 55% (42/77) of the participants thought that Boris should connect to an experience machine, we can conclude that experience machine scenarios do *not* provide evidence of widespread agreement about the relative intrinsic prudential value of reality and how our experiences feel to us on the inside. Recall that the great power of Nozick's experience machine thought experiment was that nearly everyone agreed that connecting to an experience machine was a bad choice even though *a lot* more enjoyment

³⁶ We have reason to believe that the Stranger NSQ scenario is not much more complicated than Nozick's scenario because the number of un-informational responses to each is very similar. About 14% (11/77) of participants responding to the Stranger NSQ scenario gave un-informational responses, only very slightly more than the 14% (11/77) of participants responding to Nozick's scenario.

was offered by a life connected to the machine. Many people then inferred that directly connecting with reality must be the reason for this because reality was the only obvious difference between the lives on offer. The results of the new scenarios presented here provide good evidence that the widespread agreement about Nozick's scenario was guided more by status quo bias and other irrelevant factors than it was by the value of reality. This can be seen by comparing the 18% (14/77) of respondents who provided 'reality, truth, real autonomy, or something related' as the main justification for preferring reality over an experience machine life to the 38% difference between the Stranger NSQ and Nozick's scenarios (which provides a plausible estimate of the total effect of status quo bias and other confounding factors on responses to Nozick's scenario). Therefore, contemplation of the experience machine should no longer give us a *prima facie* reason for rejecting hedonism and other internalist mental state theories of well-being.

The results of the Stranger NSQ scenario certainly provide no endorsement of hedonism and other internalist mental state theories of well-being, however. About 18% (14/77) of participants still mentioned that having a veridical connection between their experiences and the cause of those experiences matters to them ('reality, truth, real autonomy, or something related'). For hedonism or any other internalist mental state theory of well-being to be true, anything that is outside of the internal aspects of our experiences (such as whether they are, unbeknownst to us, caused by a machine) has no intrinsic value and should not matter to us over and above the positive experiences it might lead to. Contemplation of the experience machine, therefore, still produces some judgements that constitute evidence against hedonism. However, since these judgements

are now a minority, they should be considered defeasible, instead of decisive, evidence against hedonism and other internalist mental state theories of well-being. The important point is that the experience machine should no longer be considered to provide *conclusive* or even *strong* evidence that hedonism and all other internalist mental state theories of well-being are false. Of course, there may be many other arguments that do provide strong evidence that these theories are false.

Furthermore, these results should change the way that students are taught about well-being in introductory philosophy courses. The main problem is not that students are inclined away from hedonism and other internalist mental state theories of well-being (because there are several other fairly good arguments against the plausibility of such theories). Rather, the main problem is that students are taught that unrealistic thought experiments can constitute knockdown philosophical arguments without warning them about the many biases and other irrelevant factors that might be affecting our judgements about these kinds of scenarios. Indeed, instruction on how to assess the power of thought experiments as evidence for arguments should be considered a vital component of metaphilosophy and introductory philosophy courses. Since the differences in responses elicited by Nozickian, reversed, and Stranger NSQ scenarios are so large, the experience machine thought experiment (in all of its guises) would be an excellent example for this kind of instruction. Long live the experience machine (again)!³⁷

References

Attfield, R., 1987. *A Theory of Value and Obligation*. London: Croom Helm.

³⁷ Thanks to Paul Jose for very helpful advice on the statistical analysis and thanks to Nicholas Agar, Ramon Das, Simon Keller, David Gilbert, and Simon Clarke for helpful comments on earlier drafts.

- Baggini, J., & P. S. Fosl, 2007. *The Ethics Toolkit: A Compendium of Ethical Concepts and Methods*, Wiley-Blackwell.
- Becker, L. C., 1992. Good Lives: Prolegomena, in E. F. Paul, F. D. Miller, Jr., & J. Paul (eds.), *The Good Life and the Human Good*. Cambridge: Cambridge University Press: 15–37.
- Bloomfield A. N., J. A. Sager, D. M. Bartels, & D. L. Medin, 2006. Caring about Framing Effects, *Mind & Society*, 5(2): 123–138.
- Bok, S., 2010. *Exploring Happiness: From Aristotle to Brain Science*. New Haven: Yale University Press.
- Bostrom, N. & T. Ord, 2006. The Reversal Test: Eliminating Status Quo Bias in Applied Ethics, *Ethics*, 116: 656–679.
- Brink, D., 1989. *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
- Brülde, B., 2007. Happiness Theories of the Good Life, *Journal of Happiness Studies*, 8: 15–49.
- Chen M. K., V. Lakshminaryanan, & L. R. Santos, 2006. The Evolution of Our Preferences: Evidence from Capuchin Monkey Trading Behavior, *Journal of Political Economy*, 114: 517–537.
- Crandall, C. S., S. Eidelman, L. J. Skitka, & S. G. Morgan, 2009. Status Quo Framing Increases Support for Torture, *Social Influence*, 4: 1–10.
- Darwall, S., 1997. Self Interest and Self Concern, *Social Philosophy and Policy*, 14: 158–178.
- De Brigard, F., 2010. If You Like it, Does it Matter if it's Real?, *Philosophical Psychology*, 23(1): 43–57.

- Druckman, J. N., 2001. Evaluating Framing Effects, *Journal of Economic Psychology*, 22: 91–101.
- Eidelman, S., & Crandall, C. S., 2012. Bias in Favor of the Status Quo, *Social and Personality Psychology Compass*, 6(3): 270–281.
- Feldman, F., 2002. The Good Life: A Defense of Attitudinal Hedonism, *Philosophy and Phenomenological Research*, LXV(3): 604–628.
- Feldman, F., 2004. *Pleasure and the Good Life*, Oxford: Clarendon Press.
- Finnis, J., 1980. *Natural Law and Natural Rights*. Oxford: Clarendon Press.
- Finnis, J., 1983. *Fundamentals of Ethics*. Washington, DC: Georgetown University Press.
- Gilbert, D., 2006. *Stumbling on Happiness*, New York: Alfred A. Knopf.
- Gilovich, T., D. W. Griffin, & D. Kahneman, 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press.
- Griffin, J., 1986. *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford: Clarendon Press.
- Hartman, R. S., M. J. Doane, & C.-H. Woo, 1991. Consumer Rationality and the Status Quo, *The Quarterly Journal of Economics*, 106(1): 141–162.
- Hausman, D. M., 2010. Hedonism and Welfare Economics, *Economics and Philosophy*, 26: 321–344.
- Haybron, D. M., 2008. Happiness, the Self and Human Flourishing, *Utilitas*, 20(1): 21–49.
- Hooker, B., 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Clarendon.
- Hurka, T., 2011. *The Best Things in Life: A Guide to What Really Matters*. New York: Oxford University Press.
- Jarden, A. *et al.* (no date). Further Assessment Study of the International Wellbeing Study.

- Jollimore, T., 2004. Meaningless Happiness and Meaningful Suffering, *The Southern Journal of Philosophy*, 42(3): 333–347.
- Kagan, S., 1998. Normative Ethics, Boulder: Westview Press.
- Kagan, S., 2009. Well-Being as Enjoying the Good, *Philosophical Perspectives: Ethics*, 23: 253–272.
- Kahneman, D., J. L. Knetsch, & R. H. Thaler, 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias, *The Journal of Economic Perspectives*, 5(1): 193–206.
- Kahneman, D., & A. Tversky, 2000. *Choices, Values, and Frames*, Cambridge: Cambridge University Press.
- Kawall, J., 1999. The Experience Machine and Mental State Theories of Well-being, *The Journal of Value Inquiry*, 33: 381–387.
- Kazez, J., 2007. *The Weight of Things: Philosophy and the Good Life*, Wiley-Blackwell.
- Keller, S., 2009. Welfare as Success, *Noûs*, 43(4): 656–683.
- Kempf, A., & S. Ruenzi, 2006. Status Quo Bias and the Number of Alternatives: An Empirical Illustration from the Mutual Fund Industry, *Journal of Behavioral Finance*, 7: 204–213.
- Knetsch, J. L., 1989. The Endowment Effect and Evidence of Nonreversible Indifference Curves, *American Economic Review*, 79: 1277–1284.
- Knetsch, J. L., & J. A. Sinden, 1984. Willingness to Pay and Compensation Demand: Experimental Evidence of an Unexpected Disparity in Measures of Value, *Quarterly Journal of Economics*, 99: 507–521.
- Kolber, A., 1994. Mental Statism and the Experience Machine, *Bard Journal of Social Sciences*, 3(3–4): 10–17.

- Kraut, R., 2007. *What is Good and Why: The Ethics of Well-Being*, Cambridge, MA.: Harvard University Press.
- Kymlicka, W., 1990. *Contemporary Political Philosophy: An Introduction*, Oxford: Oxford University Press.
- Lieberman, M., 2000. Intuition: A Social Cognitive Neuroscience Approach, *Psychological Bulletin*, 126: 109–137.
- Myers, D. G., 2002. *Intuition: Its Powers and Perils*, New Haven: Yale University Press, 2004.
- Nickerson, R. S., 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, *Review of General Psychology*, 2: 175–220.
- Nozick, R., 1974. *Anarchy, State, and Utopia*, Oxford: Blackwell, 1991.
- Nozick, R., 1989. Happiness, in his *The Examined Life*, New York: Simon and Schuster, 99–117.
- Railton, P., 1984. Alienation, Consequentialism, and the Demands of Morality, *Philosophy & Public Affairs*, 13(2): 134–171.
- Rivera-López, E., 2007. Are Mental State Welfarism and Our Concern for Nonexperiential Goals Incompatible?, *Pacific Philosophical Quarterly*, 88: 74–91.
- Samuelson, W., & R. Zeckhauser, 1988. Status Quo Bias in Decision Making, *Journal of Risk and Uncertainty*, 1: 7–59.
- Smith, B. 2011. Can We Test the Experience Machine?, *Ethical Perspectives*, 18 (1): 29–51.
- Sobel, D., 2002. Varieties of Hedonism, *Journal of Social Philosophy*, 33(2): 240–256.
- Sumner, L., 1996: *Welfare, Happiness, & Ethics*. Oxford: Clarendon press.
- Thomson, G., 1987. *Needs*. London: Routledge and Kegan Paul.

- Tiberius, V., 2004. Cultural Differences and Philosophical Accounts of Well-Being, *The Journal of Happiness Studies*, 5: 293–314.
- Tiberius, V. & Hall, A., 2010: Normative Theory and Psychological Research: Hedonism, Eudaimonism, and Why it Matters, *Journal of Positive Psychology*, 5(3): 212–225.
- Tversky, A., & D. Kahneman, 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model, *Quarterly Journal of Economics*, 106(4): 1039–1061.
- Van Wyk, R. N., 1990. *Introduction to Ethics*, New York: St. Martin's Press.
- Weijers, D., 2011a. Hedonism, *Internet Encyclopedia of Philosophy*. Available from: <http://www.iep.utm.edu/hedonism/>
- Weijers, D., 2011b. The Experience Machine Objection to Hedonism, in *Just the Arguments*, Michael Bruce & Steven Barbone (eds.), Wiley-Blackwell, pp. 229–231.
- Weijers, D., 2011c. Reality Doesn't Really Matter, in *Inception and Philosophy*, D. Kyle Johnson (ed.), Wiley-Blackwell, pp. 92-107.
- Weijers, D., forthcoming. Intuitive Biases in Judgements about Thought Experiments: The Experience Machine Revisited, *Philosophical Writings*.
- Weijers, D., 2012. We Can Test the Experience Machine, *Ethical Perspectives*, 19(2), pp. 261-268.
- Woodward, J., & Allman, J., 2007. Moral Intuition: Its Neural Substrates and Normative Significance, *Journal of Physiology-Paris*, 101(4–6): 179–202.