

University of Groningen

Nozick's experience machine

Hindriks, Frank; Douven, Igor

Published in:
Philosophical Psychology

DOI:
[10.1080/09515089.2017.1406600](https://doi.org/10.1080/09515089.2017.1406600)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Hindriks, F., & Douven, I. (2018). Nozick's experience machine: An empirical study. *Philosophical Psychology*, 31(2), 278-298. <https://doi.org/10.1080/09515089.2017.1406600>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Nozick's experience machine: An empirical study

Frank Hindriks & Igor Douven

To cite this article: Frank Hindriks & Igor Douven (2018) Nozick's experience machine: An empirical study, *Philosophical Psychology*, 31:2, 278-298, DOI: [10.1080/09515089.2017.1406600](https://doi.org/10.1080/09515089.2017.1406600)

To link to this article: <https://doi.org/10.1080/09515089.2017.1406600>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 01 Dec 2017.



Submit your article to this journal [↗](#)



Article views: 1762



View Crossmark data [↗](#)

Nozick's experience machine: An empirical study

Frank Hindriks^a and Igor Douven^b

^aDepartment of Philosophy, University of Groningen, Groningen, Netherlands; ^bSND/CNRS, Sorbonne University, Paris, France

ABSTRACT

Many philosophers deny that happiness can be equated with pleasurable experiences. Nozick introduced an experience machine thought experiment to support the idea that happiness requires pleasurable experiences that are “in contact with reality.” In this thought experiment, people can choose to plug into a machine that induces exclusively pleasurable experiences. We test Nozick's hypothesis that people will reject this offer. We also contrast Nozick's experience machine scenario with scenarios that are less artificial, and offer options which are less invasive or disruptive than being connected to a machine, specifically scenarios in which people are offered an experience pill or a pill that improves overall functioning.

ARTICLE HISTORY

Received 5 June 2016
Accepted 23 August 2017

KEYWORDS

Experience machine;
happiness; Nozick

1. Introduction

In order to show that happiness cannot be equated with pleasurable experiences, Robert Nozick (1974) invented a thought experiment involving an experience machine. In this thought experiment, we are to imagine that we can choose to plug into an experience machine that ensures our having exclusively pleasurable experiences. Nozick (1974, p. 646) supposes that we will reject this offer, because we want to live a life that is “in contact with reality” (more on this below). This paper takes Nozick's thought experiment as a starting point for a survey study, aiming to thereby determine the level of support for the intuition this thought experiment seeks to bring into relief.

Empirical work on Nozick's “contact intuition” revealed that this intuition is not as universally shared as many in the philosophical community have supposed (De Brigard, 2010; Weijers, 2014). We probe more deeply by considering variations of Nozick's original scenario that feature interventions which are both more realistic and less invasive than hooking someone up to a machine. Specifically, in

CONTACT Frank Hindriks  f.a.hindriks@rug.nl

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Experiment 1, we also present the option of taking an experience pill that has the same effect as the experience machine, as well as the option of taking a functioning pill that improves one's overall functioning. Inspired by Nozick, we predict that the less invasive an intervention is—the less it severs contact with reality—the more people will be prepared to accept it. We contrast our findings with those of the aforementioned earlier empirical studies. In a second experiment, we consider still further variations of Nozick's scenarios, which are parallel to the materials from Experiment 1 except that instead of offering an improvement over the *status quo*, the scenarios promise to prevent a pending loss. Recent research suggests that the difference in “valence” between the scenarios from Experiment 1 and those from Experiment 2 may have an effect on people's willingness to accept the interventions on offer. We investigate whether it does by comparing the data from the two experiments.

2. Theoretical background

Hedonism is the view that happiness can be reduced to pleasurable experiences. When considering this position, Nozick (1974, p. 644) asks, “What else can matter to us, other than how our lives feel from the inside?” A hedonist will be inclined to read this as a rhetorical question, with “nothing” as the obvious answer. Nozick (1974, p. 646), however, rejects hedonism and suggests instead, “Perhaps what we desire is to live (an active verb) ourselves, in contact with reality (And this machines cannot do *for* us).” Dan Haybron (2011, p. 27) formulates Nozick's key point as follows: “Beyond having positive mental states, it seems to matter both that our lives go well and that our state of mind is appropriately related to how things are.” Nozick's point appears to be that people intrinsically value contact with reality in the sense that they value their mental states being veridical and their experiences being genuinely of our world.

Nozick mentions a number of things other than pleasurable experiences that matter to us. It matters to people that they *do* something rather than merely have experiences. Furthermore, it matters to them what kind of person they are, whether, for instance, they are courageous, kind, or intelligent. In light of this, Nozick (1974, p. 645) even concludes that “plugging into the machine is a kind of suicide.” The final consideration he mentions is that people care about being in touch with a reality that is deeper or more important than one that is manufactured, one that is real rather than artificial.

2.1. Previous research

Thus far, empirical research concerning the experience machine has focused on factors that might bias people's responses to Nozick's scenario (De Brigard, 2010; Weijers, 2014). Wayne Sumner suggests that reading that scenario might trigger a number of strictly irrelevant thoughts:

We immediately begin to imagine the ways in which things could go horribly wrong. How do we know that the technology is foolproof? What happens if there is a power failure? Suppose the operators of the machine are really sadistic thrill-seekers, or the premises are overrun by fundamentalist zealots. (1992, p. 216)

In view of these considerations, one might doubt whether the thought experiment reveals anything about what people value, rather than about the factors that bias them. As a more general concern, Sumner (1992, p. 216) mentions that

for the experience machine to yield any philosophically interesting results we must imagine ourselves in a world very different from our own—so different that any choices we make in that world might tell us very little about how we think our lives should go in the real world.

Sumner notes that the philosophical point of the experience machine heavily depends on the supposition that the biasing factors have been neutralized. Dan Weijers (2014) tries to formulate a vignette that neutralizes at least a number of such factors. The first step he takes is to test the scenario as Nozick originally presented it. This lengthy passage features “superduper neuropsychologists” who stimulate your brain and pre-program your life’s experiences; you will be “floating in a tank, with electrodes attached to your brain” (Nozick, 1974, pp. 42–43). From an extensive library that businesses have developed by researching the lives of many other people, you can choose the experiences you want to have for the next two years. Then you get some time to select experiences for another two years. It is stipulated that while you are hooked up to the machine, you will not be aware of that fact.

Weijers (2014, p. 520) presents the participants to his first study with the following question: “What is the best thing for you to do for yourself in this situation?” Participants can choose between “Permanently plug in to an Experience Machine” and “Never plug in to an Experience Machine,” and they are asked to briefly explain their choice. Of 79 participants, 13 (or 16 percent) indicate that they would connect to an experience machine. This could be taken to indicate that a vast majority of people share Nozick’s intuition. Many of the justifications people gave for not wanting to be connected to the machine concerned autonomy or being in touch with reality. These justifications fit well with the reasons Nozick presents for rejecting the offer.

However, Weijers goes on to note that only 47 percent of the justifications concern (in our terminology) Nozick’s contact intuition.¹ Other explanations subjects gave for their choices suggest that their responses were biased by a number of factors. One such factor is overactive imagination, which is exemplified by a participant writing that “the machine seems scary or unnatural.” Another biasing factor is what Weijers refers to as “imaginative resistance.” Participants put forward diverse considerations that fit under this heading, including “the machine might break down or not produce great experiences in the future,” “unpredictable or surprising experiences are better than pre-programmed ones,” and “bad experiences are required to appreciate good experiences or to develop properly.” Weijers also

classifies justifications along the lines of “I can’t because I have responsibilities to others” under imaginative resistance. He concludes:

The fact that 47% (31/66) of the participants who 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. (Weijers, 2014, p. 521)

Subsequently, Weijers tests a vignette that he takes to be affected less by confounding factors:

You have had a go in an Experience Machine before and know that they provide an unpredictable rollercoaster 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. (p. 521)

Almost all of the confounding factors mentioned above are addressed in this vignette in that it is explicitly stipulated that they play no role. Some others are addressed by the phrasing of the question:

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? (p. 521)

Now, 34 percent of the participants chose to connect to the experience machine, and only 31 percent provided an irrelevant justification for their choice.

Weijers goes on to investigate loss aversion as a potential source of bias. People who experience loss aversion assign greater value to avoiding losses than to securing gains. As it happens, such asymmetries in how people evaluate options all but disappear when it comes to evaluating the options faced by strangers. In view of this, Weijers contrasts the scenario just mentioned, in which a subject is asked what would be best for *him* or *her* to do, to a version in which the subject is asked what would be best for a *stranger* to do. In this Stranger scenario, 52 percent of the participants deem it best for the stranger to connect to the experience machine against 34 percent in the earlier scenario to which Weijers now refers as “the Self scenario.” Weijers explains this difference in terms of loss aversion, arguing that when subjects evaluate the choice faced by a stranger, they are less sensitive to risk and more impartial.

Loss aversion can also occur because the subject perceives the *status quo* as substantially less risky than the alternative presented. In Nozick’s scenario, reality is the *status quo*. Perhaps the reason why at least some people prefer the *status quo* is not because they intrinsically value reality. Instead, it might be because they are averse to the loss they might experience when taking the risk of being connected to an experience machine. To rule out or at least mitigate this bias, Weijers formulates a version of the Stranger scenario in which the stranger has

already switched a number of times between real life and a life generated by the experience machine (without being aware of this). As the stranger has already spent 50 percent of the recent past being connected to the experience machine, Weijers suggests, real life ceases to be the *status quo*. And he hypothesizes that the *status quo* bias will all but disappear when both options are framed as equally familiar. Fifty-five percent of participants now say that the stranger should choose “the Experience Machine life” rather than “real life” (Weijers, 2014, p. 526). Even though the difference between this Stranger No Status Quo (NSQ) scenario and the Stranger scenario discussed previously is not statistically significant, there is a significant difference with the Self scenario ($p = .008$, using a one-tailed Fisher’s exact test; Weijers, 2014, p. 533, note 28).

Weijers concludes that Nozick’s scenario is less useful for evaluating hedonism than it is commonly taken to be. At the same time, he is optimistic about the extent to which the scenarios he uses succeed in ruling out biases. In particular, he suggests that the Stranger NSQ scenario is “relatively bias free” (Weijers, 2014, p. 528). He goes on to argue 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” (Weijers, 2014, p. 529). The upshot is that the experience machine provides “defeasible” rather than “decisive” evidence against hedonism (p. 529).

Prior to Weijers, Felipe De Brigard (2010) defended the more radical claim that, because of the status quo bias, experimental findings concerning the experience machine are of little use. He argued that Nozick’s scenario is inadequate as a test of hedonism. In contrast to Weijers, however, De Brigard does not test Nozick’s original scenario. Instead, he uses a vignette in which subjects are to imagine that, unbeknownst to them, they have been hooked up to an experience machine for some time now. The underlying idea is that people are averse to abandoning the life they have been experiencing thus far, irrespective of whether it is virtual or real. This means that the status quo bias might not only explain the negative responses people (allegedly) tend to have toward Nozick’s original scenario, it also provides reason to expect people to respond negatively to scenarios in which they can choose to return to reality in a situation in which they have been connected to the experience machine for some time. De Brigard (2010) finds that people are more or less equally divided on the question of whether they would like to remain connected to the machine (41% of the participants wanted to disconnect). He concludes that, as in this case there is not a vast majority that rejects the option, people’s responses are influenced by some biasing factor.²

De Brigard formulates one scenario in which people are hardly divided. In this scenario, the alternative to remaining connected is living a very unhappy life, more specifically being a prisoner in a maximum security prison.³ When that is the alternative, the vast majority—87 percent—choose to remain connected to the machine (De Brigard, 2010, p. 47). This suggests that people do not value reality more than avoiding a life full of unpleasant experiences. In light of his experiments, De Brigard (2010, p. 51) concludes that “some people may prefer

to remain unplugged, not because they value reality, but because they are averse to losing their status quo.”

One might doubt that De Brigard’s findings pose a significant challenge to Nozick’s contact intuition. De Brigard (2010, p. 48) notes that the fact that people are divided in the neutral and positive conditions could be taken to mean that more than half of the participants share Nozick’s contact intuition and prefer to go back to reality in the neutral condition.⁴ This interpretation certainly fits the findings in the positive condition. In that condition, the quality of the experiences inside and outside of the machine is similar, which means that hedonists could respond both ways. However, the neutral condition seems to provide evidence against hedonism, as the machine is clearly the better option when evaluated only in terms of experiences. In the negative condition, the fact that the vast majority favors the machine does not establish that people exclusively care for pleasurable experiences. Instead, it supports the idea that avoiding unpleasant experiences matters to people, possibly as one among several things, which could include living a life in touch with reality.

The upshot is that, even though the vast majority (84 percent) rejects the option of being connected to an experience machine when presented with Nozick’s original scenario, there is some reason to believe that these responses are biased by loss aversion. Attempts to correct for this bias, for instance in the way Weijers tried to do this, result in acceptance rates up to 55 percent (59 percent when the option is to remain connected).

2.2. Plan

To decrease the chance that people’s responses are influenced by biasing factors such as imaginative resistance and overactive imagination, we used short and simple scenarios in our own studies, presented below. Weijers used Nozick’s original description of the experience machine, which contains a number of possibly distracting details (such as, for instance, the mentioned reference to businesses that have developed an extensive library by researching the lives of many other people). The vignettes that De Brigard uses are also lengthy because the situation he presents as the status quo is that of being already connected to the experience machine.⁵ Because the *status quo* in our vignettes is ordinary life, our vignettes can also be brief in this respect. We accommodate the concern that one can only experience pleasure if one also has unpleasant experiences by stipulating that the experience machine induces experiences that are almost exclusively pleasurable.⁶ In this way, we also try to succinctly exclude this further potential biasing factor.

To be sure, permitting longer vignettes gives one more opportunity to explicitly stipulate that certain factors are irrelevant, which might make subjects discount such factors. We believe, however, that in particular, imaginative resistance can be decreased by excluding features that are far removed from ordinary life. The

question of whether this is indeed the case can be answered, at least in part, by comparing our findings to those of Weijers and De Brigard, as we do below.⁷

3. Authenticity and valence

As intimated, the intuition Nozick sought to address with his experience machine thought experiment is a bit broader than the idea that we intrinsically value a life in contact with reality. Two other ideas mentioned in the context of the thought experiment are that it matters to people who they are, and that it matters to them what they do, not only what they experience. While it is not entirely straightforward to determine what Nozick had in mind when he identified these three aspects, we believe that each of the aforementioned considerations can be taken to capture an aspect of *being authentic*, or of *living an authentic life*.⁸ This notion of authenticity may be hard or even impossible to make formally precise, but that does not make it obscure. Indeed, we take it to be clear enough that authenticity is reflected in the traits persons have, in what they do, and in how genuine their experiences as well as their relations to the outside world are. We take it to be equally clear that the notion is of interest in and of itself, regardless of what Nozick had exactly in mind (Nozick exegesis is not the aim of this paper). We hypothesize that people value authenticity, and dub this hypothesis “the authenticity intuition.”

A main source of inspiration for our own study is the observation that authenticity admits of degrees. Someone can live a life that is more or less in contact with reality, can be closer or farther removed from her true self, and act to different degrees in accordance with her own values and desires.⁹ Based on this observation, we aim to take some first steps in investigating whether the authenticity intuition is shared by the folk. Do ordinary people really care about living an authentic life, or is it just philosophers falsely supposing that people care about authenticity?

Nozick’s experience machine scenario offers an excellent starting point for this investigation, precisely because it is not hard to come up with scenarios that are like it in important respects but that feature interventions which steer people less far away from normality. The experience machine is exceedingly invasive in that it completely detaches people from the life they ordinarily lead. People’s experiences can, however, be influenced in less intrusive ways. For this reason, we conducted studies featuring less invasive scenarios. In particular, rather than the option of being hooked up to a machine, people are presented with the option of taking a pill that induces pleasurable experiences. Because it is less invasive, the experience pill scenario could be said to preserve authenticity, or contact with reality, to a greater degree than the experience machine scenario does.

One might think that the very fact that subjective experiences are influenced directly makes them inauthentic. This thought led us to include a scenario in which a pill influences the subject’s overall functioning. Improved functioning is in turn likely to have a positive effect on the experiences someone has. As the manipulation is indirect, this scenario featuring a functioning pill would seem to

imperial authenticity to a still lesser degree than the one featuring an experience pill. The functioning pill can be related directly with one of Nozick's concerns, that for being a particular person, "A second reason for not plugging in [to the experience machine] is that we want to be a certain way, to be a certain sort of person" (1974, p. 645). Just like being closer to reality, being a particular person can presumably contribute to someone's happiness. This is how the experimental condition under discussion fits the overall experimental design. It is loosely inspired by the transformation machine that Nozick mentions, "which transforms us into whatever sort of person we'd like to be (compatible with our staying us)" (1974, p. 646). By and large, people remain the same person when their functioning is enhanced. It is their capacities that now function better. Thus, Nozick's characterization of what people care about supports our claim that people whose functioning is improved live a more authentic life than people whose experiences are manipulated directly.

In contrast to Weijers, we use only first-person scenarios, scenarios in which participants are to imagine that they face the choice at issue. We do not question that in general people are more averse to risk when they evaluate options from their own perspective as compared to options faced by a stranger. However, we doubt that a scenario featuring a stranger can capture a concern for authenticity. Authenticity, it seems to us, is primarily a first-person concern.

A second source of inspiration for our study consists in a wide range of valence asymmetries discovered in experimental philosophy.¹⁰ As it turns out, the way in which people apply apparently non-normative notions such as intentional action, freedom, and happiness is sensitive to normative factors. In particular, they turn out to be sensitive to whether or not the situation experienced or affected is good or bad. For instance, Joshua Knobe (2003) finds that people qualify the behavior of the chairman of a company whose business strategy happens to affect the environment as unintentional when this effect is beneficial, and as intentional when it is harmful.¹¹

Valence also enters De Brigard's investigation of how people respond to scenarios in which they are to imagine themselves having been connected to an experience machine. He compares three versions that differ with respect to the life that awaits them outside of the machine. That life might be good, neutral, or bad. In other words, it might have a positive, neutral, or negative valence. De Brigard finds that 87 percent of the participants want to remain connected in the negative condition, against 50 percent in the positive condition, and 46 percent in the neutral condition. This suggests that the negative valence of the outcome makes people accept the option of being hooked up to the experience machine. De Brigard takes this to be due to people's aversion to loss. We test whether valence matters in this way when the *status quo* is the ordinary life a subject leads in which he or she is neither very happy nor very unhappy. The alternative is a life in which they have almost exclusively unpleasant experiences. People can avoid this future by being connected to an experience machine (or by taking an experience pill or a functioning pill). As this pending future has a negative valence, loss aversion

leads us to also expect that when real life is the alternative, more people will want to connect to the experience machine (or take one of the pills).

4. Research questions

The authenticity intuition can be tested by presenting people with a vignette that features Nozick's experience machine. For the reasons explained previously, we used a version of Nozick's scenario stripped down to its essence. Our first research question (Q1) is whether, using such a version, we can replicate Weijers' finding that only a small percentage of participants (16 percent, in his study) accepts the offer of being hooked up to the experience machine.

A stronger test of the same intuition can be developed once it is recognized that authenticity is a matter of degree. The experience machine is rather invasive in that it completely disconnects people from reality. By considering less invasive manipulations of people's experiences, we can test whether people become more accepting of the envisaged scenario. To this end, we formulated a vignette that features an experience pill that has the same effect as the experience machine. In contrast to people who are hooked up to an experience machine, people who take an experience pill still live a life in their ordinary surroundings, meaning, among other things, that they have genuine personal relations (see note 8). Given that people taking the pill would still live in contact with reality in the sense that their mental states would be veridical and their experiences genuinely of our world, we expect that people will be more inclined to take the experience pill than to connect to the experience machine. Whether we are right to expect this is our second research question (Q2).

Nozick's remark regarding whether people care about the kind of persons they are suggests that part of what might be problematic about the experience machine is that their personality is not reflected in the kind of experiences they have. To mitigate this feature of the original thought experiment, we prompt people to imagine that, rather than their subjective experiences, the way they *function* is enhanced by means of a pill. In other words, by taking the functioning pill, people's overall functioning improves substantially. As it is *their* capabilities that are enhanced or *their* way of functioning, the experiences they will have are more authentic than the ones that are directly induced. In the extreme case, people might see little or no reason to discount the resulting experiences as artificial. Indeed, it could be argued that we live our lives most authentically when we can be our best possible selves. This leads us to expect that more people will choose the functioning pill than the experience pill. Whether this is correct is the third research question (Q3).

Underlying our expectations is the thought that being hooked up to an experience machine is a major intervention that completely disconnects an agent from reality. Taking a pill rather than being hooked up to a machine is less invasive. The infringement on authenticity is even less (or might be deemed to be completely

absent) when the pill influences people's functioning rather than their experiences directly.

Our fourth and final research question (Q4) concerns the valence of the outcomes. This question is inspired by the findings concerning loss aversion discussed in Section 2. More indirectly, it is also motivated by many recent empirical findings within experimental philosophy that reveal the significance of normative factors for apparently normatively neutral concepts (such as intentional action, causation, freedom, and happiness). Although the interpretation of these findings is a matter of ongoing controversy, they suggest that normative factors such as moral valence, norm violations, normative reasons, or responsibility attributions influence our intuitions concerning concepts of the designated kind.

To investigate the role of valence in the context of Nozick's thought experiment, we formulated vignettes in which, rather than inducing pleasurable experiences or enhancing functioning (the positive condition), a machine or a pill prevents people from having negative experiences or from their functioning deteriorating (the negative condition). We hypothesize that people will be more inclined to accept measures that prevent a bad outcome as compared to induce a good outcome, or differently put, that because people may suffer from loss aversion (De Brigard, 2010; Weijers, 2014), they will tend to care less about living in contact with reality when it comes to preventing unhappiness as compared to bringing about happiness.

5. Experiment 1

This experiment was designed to investigate questions Q1–Q3.

5.1. Participants

There were 249 participants in this experiment. They were recruited via CrowdFlower, which directed them to the Qualtrics platform on which the experiment was run. The participants were financially compensated for their time and effort. Repeat participation was prevented.

We first excluded participants who returned incomplete response sets, non-native speakers of English, and participants who indicated that they had not responded seriously (see Aust, Diedenhofen, Ullrich, & Musch, 2013). This left us with 224 participants. From those, we removed the fastest 2.5 percent of responders, as well as the slowest 2.5 percent, which left us with 210 participants for the final analysis ($M_{\text{age}} = 38$, $SD_{\text{age}} = 13$; 120 females).¹²

5.2. Materials and procedure

Participants were divided into six groups, each group receiving a different question. The questions concerned either the possibility of being connected to a

machine that guarantees pleasurable experiences or the possibility of taking a pill, where the pill either guarantees pleasurable experiences or improves functioning. Specifically, each of the following three vignettes was offered to two groups:

5.2.1. *Machine*

Imagine that you are presented with a choice to plug into an experience machine. Due to the machine, the experiences you will have for the rest of your life are almost exclusively pleasurable. If you accept this option, you will be permanently floating in a tank with electrodes attached to your brain (without being aware of this).

5.2.2. *Experience pill*

Imagine that you are presented with a choice to start taking an experience pill. Due to the pill, the experiences you will have for the rest of your life are almost exclusively pleasurable. If you accept this option, you will be taking the pill permanently on a daily basis (The pill will have no detrimental long-term effects on your health).

5.2.3. *Functioning pill*

Imagine that you are presented with a choice to start taking a pill that affects the way you function. Due to the pill, the way in which you function in a wide range of respects will substantially improve for the rest of your life. Specifically, your physical, cognitive, and social functioning will improve (e.g., you will become healthier, smarter, wittier, and more social). If you accept this option, you will be taking the pill permanently on a daily basis (The pill will have no detrimental long-term effects on your health).

We refer to these vignettes as MP, EP, and FP, respectively (The *P* stands for “positive,” because the vignettes refer to positive experiences and later on they will be contrasted with vignettes that refer to negative experiences, which will be used in Experiment 2).

For control purposes, we elicited responses in two different ways: One of the two groups that received a given vignette was asked for a yes/no answer—“Would you choose to plug into the experience machine/to start taking the pill?”—while the other group was asked to respond on a Likert scale from 1 = very unfavorable to 7 = very favorable, the exact question being, “What attitude do you have toward plugging into the experience machine/starting taking this pill?”

5.3. *Results*

Question Q1 asked whether Weijers’ finding that the vast majority of his participants rejected the offer of being hooked up to the experience machine could be replicated given a version of Nozick’s scenarios stripped down to its essentials. The data bearing on this question came from the group of 35 participants who had

received the MP vignette followed by the yes/no question. Of these, 10 answered positively. This is a considerably higher percentage (29 percent) than was found in Weijers (2014) study, in which 16 percent accepted the offer of being hooked up to the machine (De Brigard's, 2010 finding was exactly the same as Weijers'). A binomial test showed that this difference in percentage was significant, $p = .043$ (one-sided).¹³ On the other hand, it is still the case that a vast majority rejected the offer of being connected to the machine.

To answer questions Q2 and Q3, we started by comparing the response frequencies to the three yes/no questions. These are displayed in Figure 1. This figure shows that the proportion of positive responses to the question concerning the functioning pill is very high and (much) higher than the corresponding proportions for the other two conditions, and further that the proportion of positive responses to the question concerning the experience pill is considerably higher than the proportion of positive responses to the question concerning the machine. To be exact, the percentage of positive responses in the EP condition was 53 and that in the FP condition 89. A chi-square test of independence confirmed that the association between condition (MP/EP/FP) and response (yes/no) is highly significant and very strong: $\chi^2(2) = 25.97$, $p < .0001$; Cramér's $V = .50$. We followed up the chi-square test with a series of pairwise Fisher's exact tests, which revealed a highly significant difference between the MP and FP conditions ($p < .0001$) and a marginally significant difference between the MP and EP conditions ($p = .054$); the difference between the EP and FP conditions also came out significant ($p = .002$; all reported p values are Bonferroni–Holm adjusted).¹⁴

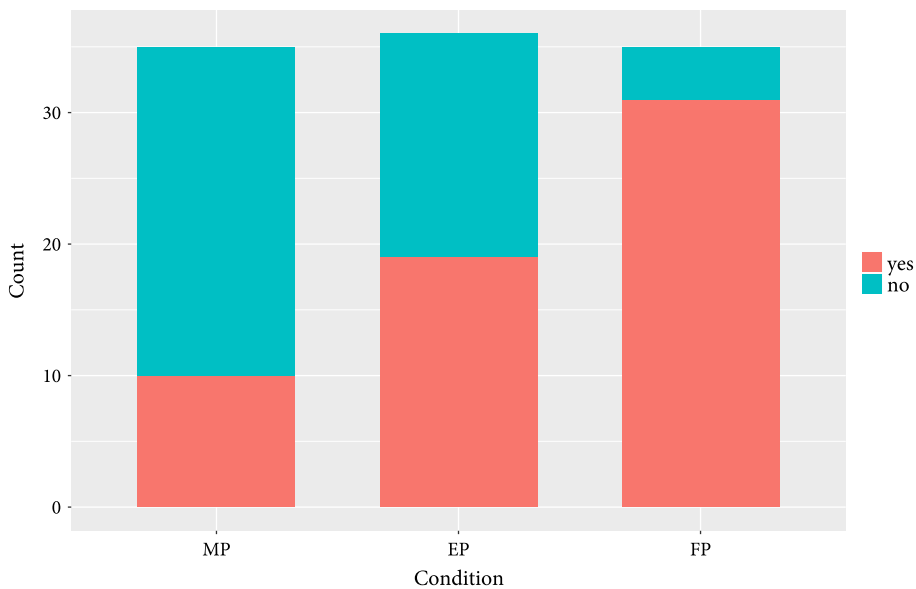


Figure 1. Yes/no responses for the three vignettes used in Experiment 1.

We obtained further information relevant to answering Q2 and Q3 from the responses to the three Likert scale questions. An overview of these responses is given in Figure 2. Visually, the trend in those responses is not too different from to the trend in the responses to the yes/no questions, with participants being more favorably inclined toward the EP and FP options than to the MP option, and also seemingly somewhat more favorably inclined to the FP option than to the EP option. Here, this impression was confirmed by a one-way ANOVA, with condition (MP/EP/FP) as independent variable and rating (from very unfavorable to very favorable) as dependent variable: $F(2, 101) = 28.05, p < .0001$; there was a very large effect of condition on ratings: $\eta_p^2 = .36$. *Post-hoc* comparisons using Tukey's HSD indicated that both the difference between the mean rating for the MP condition ($M_{MP} = 2.59, SD_{MP} = 1.89, N_{MP} = 34$) and that for the EP condition ($M_{EP} = 5.09, SD_{EP} = 1.96, N_{EP} = 35$) and the difference between the former mean rating and the mean rating for the FP condition ($M_{FP} = 5.71, SD_{FP} = 1.62, N_{FP} = 35$) were significant: both $ps < .0001$; there was no significant difference between the mean ratings for the EP and FP conditions: $p = .326$.¹⁵

5.4. Discussion

We presented participants with Nozick's experience machine scenario in a bare form. This made close to 30 percent of the participants prefer the option, which is significantly more than the percentages found in some previous research using more detail-rich versions of the experience machine scenario, but it is a minority nonetheless. So, the answer to Q1 is essentially positive: while we did not strictly

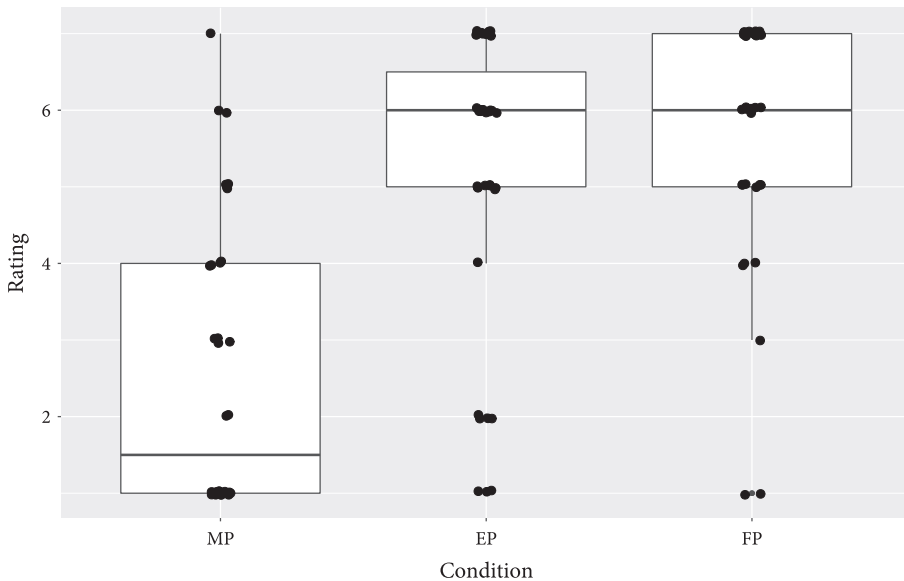


Figure 2. Likert scale responses for the three vignettes used in Experiment 1. (Individual responses are displayed with jitter to enhance their visibility.)

obtain a replication of Weijers' result for our version of the experience machine scenario, our results confirm his general finding that a majority of people appear to oppose being connected to the machine.

Furthermore, our data—both the yes/no responses and the Likert scale responses—warrant an unequivocally positive answer to Q2. Participants appeared significantly more inclined to accept the offer to take a pill, whether one guaranteeing pleasurable experiences or one guaranteeing improved functioning, than to accept the offer to be hooked up to the experience machine. The answer to Q3 is positive as well, at least as far as the yes/no responses go: significantly more participants were willing to take a pill guaranteeing improved functioning than one guaranteeing pleasurable experiences. As for the Likert scale responses, the mean response for taking the functioning pill was higher than that for taking the experience pill, but the difference was not significant. In fact, both means were rather close to the maximal rating.

As stated earlier, someone taking a pill guaranteeing pleasurable experiences could be said to be more in contact with reality than someone who is connected to a machine guaranteeing pleasurable experiences, while someone taking a pill that improves functioning could be said to be in still closer contact with reality. Thus, we predicted positive answers to Q2 and Q3, based on what we called “the authenticity intuition.” Accordingly, that we found a positive answer to Q2 and also a mostly positive answer to Q3 supports that intuition.

6. Experiment 2

The second experiment was designed to address Q4. It was basically a rerun of Experiment 1, except that all vignettes had opposite valence, in that the options offered in them prevented negative consequences, rather than generated positive consequences.

6.1. Participants

In this experiment, there were 256 participants. Recruitment, testing, and financial compensation of participants was as in Experiment 1. We also used the same selection criteria. This left us with 231 participants for the final analysis ($M_{\text{age}} = 40$, $SD_{\text{age}} = 13$; 134 females).

6.2. Materials and procedure

The materials and procedure were as in Experiment 1, the only difference being that in the machine and experience pill vignettes the participants were told that the machine/pill would prevent them from a pending future in which they would experience almost exclusively unpleasant experiences (instead of guaranteeing almost exclusively pleasurable experiences), while in the functioning pill vignette

participants were told that the pill would prevent them from a pending future in which the way they function in a wide range of respects would substantially deteriorate (instead of guaranteeing improved functioning). Because of their negative valence, we refer to these variant vignettes as MN, EN, and FN, respectively (See the Supplementary Information for the full vignettes).

6.3. Results

Figure 3 gives an overview of the responses from the three groups that were asked yes/no questions. When this figure is compared with Figure 1, only one difference stands out: in the FN condition, acceptance of the offer to take the pill is, while still the majority response, not at the same high level as in the FP condition from Experiment 1. Specifically, in the MN condition, only 26 percent gave a positive answer (29 percent in the MP condition from Experiment 1), while in the EN and FN conditions the percentages were 55 and 64, respectively, the corresponding percentages from Experiment 1 being 53 and 89.

A series of χ^2 tests showed that, while there is no significant difference between the results from the MP and MN conditions, nor between the results from the EP and EN conditions, the difference between the results from the FP and FN conditions is significant: $\chi^2(1) = 4.66, p = .031$, Cramér's $V = .29$. A χ^2 test for the results from Experiment 2 showed that there was again a significant association between condition (MN/EN/FN) and response (yes/no): $\chi^2(2) = 11.38, p = .003$, Cramér's $V = .32$. Following up this finding by pairwise Fisher's exact tests revealed that the proportion of yes/no responses in the MN condition differed significantly

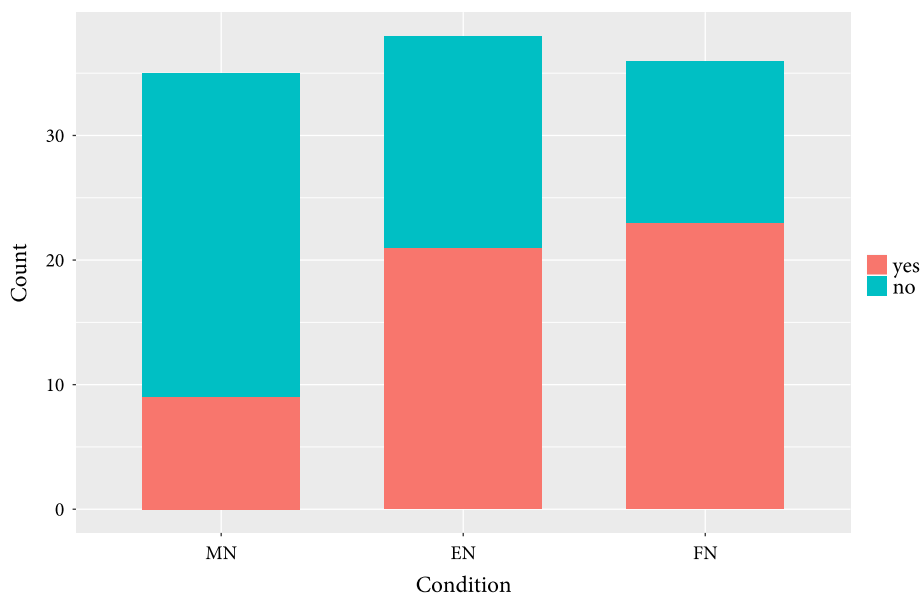


Figure 3. Yes/no responses for the three vignettes used in Experiment 2.

from both the proportion of yes/no responses in the EN condition and the proportion of yes/no responses in the FN condition ($p = .025$ and $p = .006$, respectively, these being Bonferroni–Holm adjusted p -values); there was, as expected, no significant difference between the proportions of yes/no responses in the EN and FN conditions.¹⁶

The results from the Likert scale questions are represented in Figure 4. As a comparison of this figure with Figure 2 already suggests, the differences between the conditions here were smaller than the differences between the corresponding conditions in Experiment 1: ($M_{MN} = 2.54$, $S_{MN} = 1.83$, $N_{MN} = 37$, $M_{EN} = 4.58$, $SD_{EN} = 2.10$, $N_{EN} = 38$, $M_{FN} = 4.70$, $SD_{FN} = 1.91$, $N_{FN} = 33$). Nevertheless, a one-way ANOVA with condition (MN/EN/FN) as independent variable and rating (from very unfavorable to very favorable) as dependent variable turned out significant: $F(2,105) = 13.97$, $p < .0001$, $\eta_p^2 = .210$.¹⁷

Most relevantly to Q4, we pooled the Likert scale data from Experiments 1 and 2 and conducted a 3×2 NOVA with condition (M/E/F) and valence (P/N) as between-subjects factors and response (yes/no) as dependent variable.¹⁸ This revealed a main effect of condition: $F(2, 206) = 40.28$, $p < .0001$, $\eta_p^2 = .279$. However, that was not surprising, given the outcomes of the separate one-way ANOVAs for Experiments 1 and 2, and it also has no direct bearing on Q4. Relevant to that question is the fact that there was also a main effect of valence, $F(1, 206) = 3.94$, $p = .049$ although the effect was small: $\eta_p^2 = .019$. There was no significant interaction between the factors: $F(2, 206) = 1.14$, $p = .323$, $\eta_p^2 = .011$.¹⁹

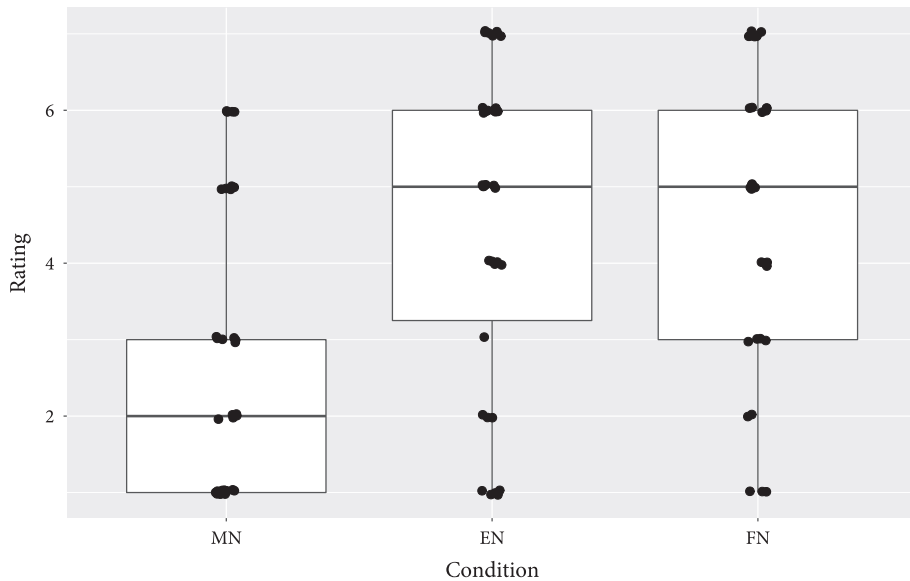


Figure 4. Likert scale responses for the three vignettes used in Experiment 2.

6.4. Discussion

Based on previous literature, there was reason to reckon with the possibility that (what we called) the valence of the vignettes might make a difference to how participants judge them. Most relevantly, in Section 1.1 we mentioned De Brigard's scenario that offers participants, as an alternative to remaining connected to Nozick's experience machine, the option of living a very unhappy life as a prisoner in a maximum security prison, an alternative that was rejected by 87 percent of his participants (they chose to remain connected to the machine). This can be explained in terms of loss aversion, and it would lead one to expect a significant difference between the P- and the N-scenarios. However, this expectation was only weakly confirmed, in that we found no more than a small effect of valence, which moreover was due to only one scenario (the functioning pill scenario).

While the main focus of Experiment 2 was Q4, it is to be noted that the results from this experiment are very similar to those from Experiment 1. In particular, we found again that a majority of participants in the experience machine condition *rejected* the offer to be connected to the machine while both in the experience pill and in the functioning pill condition, the majority of participants *accepted* the offer to take the pill. Also, the trend was again that the percentage of positive responses was highest for the functioning pill vignette, lower for the experience pill vignette, and still lower for the experience machine vignette. Thereby, the current results shed further light on questions Q1–Q3 as well, and more broadly offer further support for the authenticity intuition.

7. Conclusion

There has been much discussion in the philosophical literature about Nozick's experience machine thought experiment. We took that thought experiment, together with some of Nozick's commentary on it, as a starting point for investigating what we called "the authenticity intuition": the idea that people value living their lives in contact with reality, and also that they care about who they are and what they do. We investigated this intuition by means of two experiments. The design of these experiments was primarily motivated by the thought that authenticity comes in degrees: we can be more or less in contact with reality, stay more or less close to who we actually are, and feel to a greater or lesser degree that our actions are genuinely our own. Our materials consisted of vignettes that featured different degrees of detachment from reality and self, and were meant to establish whether those differences were reflected in our participants' acceptance rates or their attitudes toward the situations described in the vignettes.

There were two secondary motivations. First, earlier work on the experience machine had cast some doubt on the usefulness of this scenario and had pointed to various possible confounds, mostly in the form of cognitive biases that might skew people's reactions to the scenario. Our aim was to present a simple version of

Nozick's scenario that would keep such biases at bay, as much as possible. Second, it has been recognized in unrelated literature that valence—bringing about something good vs. preventing something bad from happening—can influence people's value judgments in sometimes unexpected ways. We wondered whether valence also mattered to how people valued authenticity.

The results we obtained supported earlier findings, in that participants were in majority disinclined to accept the offer of being connected to the experience machine, or, in the Likert scale conditions, held largely unfavorable attitudes toward that option. We also found that participants were more likely to accept the offer of taking an experience pill and still more likely to accept the offer of taking a pill that would improve their functioning in important ways. (Although not all differences were significant here, the trend was clear and consistent across relevant conditions.) Because these scenarios were less (the experience pill scenario) to much less (the functioning pill scenario) disruptive than Nozick's, in the sense that the result of the featured intervention would move people away from normality less than connecting to the experience machine would, these findings corroborated precisely what had been predicted on the basis of our main hypothesis (the authenticity intuition).

It was also encouraging to find that the answers to questions Q1–Q3, which concerned the authenticity intuition most directly, were the same in both experiments, meaning that the conclusions we reached about that intuition are robust and largely independent of the valence of the interventions proposed in our vignettes. There was one small effect of valence, which was due to one specific scenario in one specific condition (the functioning pill scenario in the yes/no condition). Whether this finding is a coincidence or whether it points to something deeper is an issue we leave as an avenue for future research.²⁰

Supplementary material

All supplementary information, as well as all data and the script for the statistical analyses, is available at https://osf.io/nrx2x/?view_only=d001aeff6ea24503b7e844e05f71419f.

Notes

1. Similarly, De Brigard (2010, p. 45) claims that 60 percent of the explanations his participants gave for rejecting the experience machine “had nothing to do with their preference for a real life over a virtual one.”
2. The underlying premise, which De Brigard does not test, is that the vast majority will reject being connected to the machine in Nozick's original scenario.
3. In the positive scenario, people read that their real life is that of “a multimillionaire artist living in Monaco” (De Brigard, 2010, p. 47). The neutral condition says nothing beyond “you can go back to your real life” (p. 47).
4. In response, De Brigard points out that in the neutral condition, people know nothing about what their real life will be like. He goes on to test a vignette that ends with, “your life outside is not at all like the life you have experienced so far” (De Brigard,

- 2010, p. 49). Now, 59 percent of the participants want to remain connected (against 54 percent earlier).
5. De Brigard (2010, p. 47): “It is Saturday morning and you are planning to stay in bed for at least another hour when all of the sudden you hear the doorbell. Grudgingly, you step out of bed to go open the door. At the other side there is a tall man, with a black jacket and sunglasses, who introduces himself as Mr. Smith. He claims to have vital information that concerns you directly. Mildly troubled but still curious, you let him in. ‘I am afraid I have some disturbing news to communicate to you’ says Mr. Smith. ‘There has been a terrible mistake. Your brain has been plugged by error into an experience machine created by superduper neurophysiologists. All the experiences you have had so far are nothing but the product of a computer program designed to provide you with pleasurable experiences.’”
 6. Recall that one of the participants of Weijers’ experiment rejected the option of being connected to the experience machine, because “bad experiences are required to appreciate good experiences or to develop properly” (Weijers, 2014, p. 520). De Brigard (2010, p. 47) seems to accommodate some such concerns by including the following in his vignettes: “All the unpleasantness you may have felt during your life is just an experiential preface conducive toward a greater pleasure (e.g., like when you had to wait in that long line to get tickets for that concert, remember?).”
 7. Weijers (2013) argues that including less background information does not necessarily prevent people from considering irrelevant features. People might match the scenario they read to the closest real life experiences, which include many features that are absent from the thought experiment. However, he also notes that “most philosophical thought experiments stipulate features that are so unrealistic that we have not experienced anything like them or we have experienced the very opposite of them” (Weijers, 2013, p. 22). At least some of the scenarios that we used in our studies are actually quite close to real-life situations (e.g., taking a pill that improves functioning, and certainly one that prevents deterioration, is not so different from what we hope that taking vitamin pills and other supplements does for us).
 8. Griffin (1986), p. 19, arguing against hedonism in much the same way as Nozick does, says, “ what affects well-being can only be what enters experience, and the trouble is that some of the things that persons value greatly do not. My truly having close and authentic personal relations is not the kind of thing that can enter my experience; all that can enter is what is common to both my having such relations and my merely believing that I do. And this seems to distort the nature of these values.”
 9. Note that this is not to say that authenticity is a measurable quantity. But although love is (probably) not a measurable quantity, it still makes sense to say that Jim loves Harriet more than he loved Susan. All we need to assume for the purposes of this paper is that similar comparative statements make sense for authenticity.
 10. See Knobe (2010) for an overview. Knobe argues that a unified explanation can be provided of all valence asymmetries; see Hindriks (2014) for a critical discussion.
 11. See also Phillips, Misenheimer, and Knobe (2011), who find that people discount reports of happiness when they live an objectively bad life, but not reports of unhappiness when they live an objectively good life. This suggests that unpleasant experiences suffice for unhappiness, while pleasurable experiences do not suffice for genuine happiness and less subjective values have to be present as well.
 12. In choosing these cut-off points for slowest and fastest responders, we are following the practice of other experimenters who have used the same cut-off points for the purpose of enhancing the quality of data from online surveys.

13. Using the package for Morey & Rouder (2015), we also conducted a Bayesian proportion test, obtaining a Bayes factor of 3.18 against the null hypothesis. This indicates that the data are 3.18 times more likely assuming the hypothesis that the probability of a positive response is greater than .16 than assuming the null hypothesis. According to Jeffreys (1998, p. 432) classification scheme, this means that our data substantially support the alternative hypothesis.
14. Again, we obtained basically the same results conducting a Bayesian analysis using the package, which offers an implementation of Günel and Dickey's (1974) contingency table Bayes factor test. Specifically, we found a Bayes factor of 67×10^3 in favor of an association between condition (MP/EP/FP) and response (yes/no), and conducting separate contingency table tests for pairs of levels of the condition, we found a Bayes factor of 2.34 in favor of an association between condition (MP/EP) and response (yes/no), a Bayes factor of 231×10^3 in favor of an association between condition (MP/FP) and response (yes/no), and a Bayes factor of 65.72 in favor of an association between condition (EP/FP) and response (yes/no).
15. A series of Bayesian ANOVAs, conducted using the package again, and all of which had rating as response variable, yielded very similar results. We obtained a Bayes factor of 45×10^6 favoring a model with condition (MP/EP/FP) as predictor over the intercept only model, a Bayes factor of 13×10^2 favoring a model with condition (MP/EP) as predictor over the intercept only model, a Bayes factor of 24×10^6 favoring a model with condition (MP/FP) as predictor over the intercept only model, and a Bayes factor of 1.64 favoring the intercept only model over a model with condition (EP/FP) as predictor.
16. We obtained similar results from a series of contingency table Bayes factor tests, like the ones reported in note 14. We found a Bayes factor of 21.42 in favor of an association between condition (MN/EN/FN) and response (yes/no), a Bayes factor of 7.23 in favor of an association between condition (MN/EN) and response (yes/no), a Bayes factor of 52.50 in favor of an association between condition (MN/FN) and response (yes/no), and a Bayes factor of 2.73 in favor of the null hypothesis of no association between condition (EN/FN) and response (yes/no).
17. Here, the Bayesian analysis concurred as well. A Bayesian ANOVA with condition (MN/EN/FN) as independent variable and rating as dependent variable yielded a Bayes factor of 4400.08 favoring the model with the condition as independent variable over the intercept only model. Furthermore, we found a Bayes factor of 684.13 favoring a model with condition (MN/EN) as predictor over the intercept only model, a Bayes factor of 1940.77 favoring a model with condition (MN/FN) as predictor over the intercept only model, and a Bayes factor of 3.98 favoring the intercept only model over a model with condition (EN/FN) as predictor.
18. M is a label for the union of the responses in the MP and MN conditions in Experiments 1 and 2, respectively, and similarly for E and F. P is a label for the responses from Experiment 1 (in which the vignettes had positive valence) and N is a label for the responses from Experiment 2 (in which the vignettes had negative valence).
19. A Bayesian ANOVA yielded a Bayes factor of 2.5×10^{12} favoring the model with condition as independent variable over the intercept only model, but a Bayes factor of 1.35 favoring the intercept only model over the model with valence as independent variable, a Bayes factor of 2.4×10^{12} favoring the intercept only model over the model with both condition and valence as independent variables, and a Bayes factor of 5×10^{11} favoring the model with condition, valence, as well as their interaction as independent variables over the intercept only model.

20. We are greatly indebted to three anonymous referees for valuable comments on previous versions of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, 45(2), 527–535.
- De Brigard, F. (2010). If you like it, does it matter if it's real? *Philosophical Psychology*, 23(1), 43–57.
- Griffin, J. (1986). *Well-being: Its meaning, measurement and moral importance*. New York, NY: Oxford University Press. Retrieved from <http://www.oxfordscholarship.com/view/10.1093/0198248431.001.0001/acprof-9780198248439>
- Günel, E., & Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, 61, 545–557.
- Haybron, D. (2011). Happiness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2011 Edition). Retrieved from <http://plato.stanford.edu/archives/fall2011/entries/happiness/>
- Hindriks, F. (2014). Normativity in action: How to explain the Knobe effect and its relatives. *Mind & Language*, 29, 51–72.
- Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 190–194.
- Knobe, J. (2010). The person as moralist account and its alternatives. *Behavioral and Brain Sciences*, 33(04), 353–365.
- Morey & Rouder. (2015). *Computation of Bayes factors for common designs*, package version 0.9.11-1. Retrieve from <http://CRAN.R-project.org/package=BayesFactor>
- Nozick, R. (1974). *Anarchy, state, and utopia*. New York, NY: Basic Books.
- Phillips, J., Misenheimer, L., & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review*, 3, 320–322.
- Sumner, L. W. (1992). Welfare, happiness, and pleasure. *Utilitas*, 4(02), 199–223.
- Weijers, D. (2013). Intuitive biases in judgments about thought experiments: The experience machine revisited. *Philosophical Writings*, 41, 17–31.
- Weijers, D. (2014). Nozick's experience machine is dead, long live the experience machine!. *Philosophical Psychology*, 27(4), 513–535.