

**Nozick's experience
machine thought
experiment has
many problems, but
the freebie problem
ain't one of 'em**

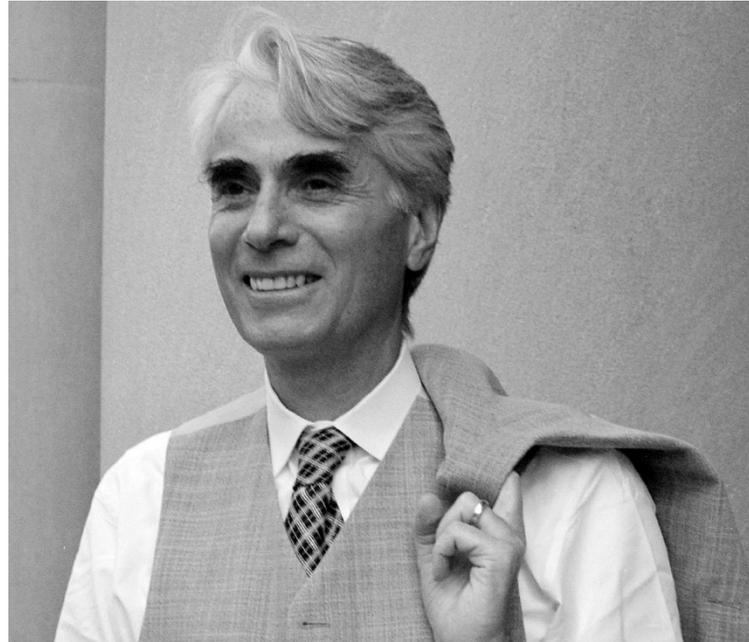
Some things I'll try to do

1. Background I: Philosophy
2. Nozick's experience machine as an argument against hedonism
3. Recent criticisms of 2
4. Eden Lin's "The experientially identical lifetime comparison argument" version (a response to 3)
5. Background II: Psychology
6. Explain the "freebie problem"
7. Point out that Lin's new version suffers from the problem
8. Conclude that the debate should go back to point 3

1. Background I: Philosophy

- Methodology in moral theory and wellbeing
 - We need intuitions about value claims to make any positive argument and many negative ones
 - Cases tend to trump theories/principles/rationales
- Wellbeing
 - The prudential good life
 - What ultimately makes a person's life go well for them?
- Prudential Hedonism
 - All and only pleasure is directly good for us, all and only pain is bad for us
- Internalist Quantitative Prudential Hedonism
 - Only the duration and intensity of how our experiences feel to us on the inside ultimately matters for wellbeing
- Experience machine cases
 - Which life is better?: real life vs. much happier machine-generated life

2a. Nozick's experience machine



- Nozick, R. (1974). “The Experience Machine,” from his *Anarchy, State and Utopia*. Basic Books, pp. 42-45.

2b. Some experience machine arguments against hedonism

1. No one would connect to an experience machine
 2. An actual hedonist would connect to an experience machine
 3. Therefore, there are no actual hedonists
-
1. If hedonism were true, the vast majority of reasonable people would connect to an experience machine (in Nozick's set up)
 2. The vast majority of reasonable people would not connect to an experience machine (in Nozick's set up)
 3. Therefore, hedonism is false

2c. The best(?) experience machine argument against hedonism

- AP1. In terms of the internal aspects of our experiences, an experience machine life would be much better than a life in reality. (Stipulated in thought experiment)
- AP2. When instructed to ignore their responsibilities to others, the vast majority of reasonable people report preferring reality over a life in an experience machine. (Empirical claim)
- AP3. The best explanation for AP2 is that reality matters intrinsically to the vast majority of reasonable people
- AP4. Inference to the best explanation: If a hypothesis is the best explanation of an observation, then it is rational to believe that hypothesis is true. (Standard methodological premise)
 - AC1. Therefore, it is rational to believe that reality matters intrinsically to the vast majority of reasonable people. (Modus ponens AP3, AP4)
- AP5. The best explanation for reality mattering intrinsically to the vast majority of reasonable people is that reality has intrinsic prudential value
- AP6. Inference to the best explanation. (Standard methodological premise)
 - AC2. Therefore, it is rational to believe that reality has intrinsic prudential value. (Modus ponens AP5, AP6)
- AP7. If internalist prudential hedonism is true, then the internal aspects of pleasure and pain are the only things of intrinsic prudential value (or disvalue) in a life. (Stipulated definition)
 - AC3. Therefore, it is rational to believe that internalist prudential hedonism is false. (Modus tollens, AC2, AP7)

Weijers, Dan & Schouten, Vanessa (2013). An assessment of recent responses to the experience machine objection to hedonism, *Journal of Value Inquiry*, 47(4), 461-482.

2d. The point of the best(?) experience machine argument against hedonism

AP2. When instructed to ignore their responsibilities to others, the vast majority of reasonable people report preferring reality over a life in an experience machine. (Empirical claim)

AP3. The best explanation for AP2 is that reality matters intrinsically to the vast majority of reasonable people

- For the argument to work, the best explanation for experience machine judgments needs to be the that people care about the purported value
- People can care about the wrong things, but the methods of ethics requires intuitions about value at some stage
 - The *vast majority of reasonable people* is an arbitrary but useful standard

3a. Recent criticisms of the experience machine argument

- Lots of criticisms of what people would or should do when faced with the experience machine choice
- E.g., the machine is too scary, so people will avoid it for an irrelevant reason
- E.g., I love my partner and kids, so I wouldn't connect
- For a review, see: Weijers, Dan & Schouten, Vanessa (2013). An assessment of recent responses to the experience machine objection to hedonism, *Journal of Value Inquiry*, 47(4), 461-482.

3b. Empirical criticisms of the experience machine argument

AP2. When instructed to ignore their responsibilities to others, the vast majority of reasonable people report preferring reality over a life in an experience machine. (Empirical claim)

AP3. The best explanation for AP2 is that reality matters intrinsically to the vast majority of reasonable people

- AP3 is false. Flipping or trying to diminish the status quo massively changes peoples judgments (even though it was not thought relevant)
- AP2 is false when “clean” scenarios are used
- De Brigard, Felipe. (2010). If you like it, does it matter if it's real? *Philosophical Psychology*, 23(1), 43-57.
- Weijers, Dan. (2014). Nozick's experience machine is dead, long live the experience machine!. *Philosophical Psychology*, 27(4), 513-535.

4a. Eden Lin's "experientially identical lifetime comparison argument"

- Scenario: A and B have experientially identical lives, but B is in an experience machine (and has no people that care about or depend on him)
 1. A is higher in welfare than B
 2. If hedonism is true, then A is not higher in welfare than B
 3. Therefore: Hedonism is false
- Lin, Eden. (2016). How to use the experience machine, *Utilitas*, 28(3): 314-332.

4b. Commentary on the “experientially identical lifetime comparison argument”

- Lin claims that his argument avoids the biases/problems of Nozick’s:
- His argument is “not susceptible to the status quo bias”, since neither A nor B changes (p. 322)
- His argument “doesn’t rely on any claims about whether we would or should plug in... Nor does it rely on any claims about what we desire or value.” (p. 322)
- But, how do we know P1 (“A is higher in welfare than B”) is true?
 - Value judgments (+experiment: 90%+ Rowland, R. (2017). Our Intuitions About the Experience Machine. *J. Ethics & Soc. Phil.*, 12, 110.)
- What if I told you those judgements were biased in such a way that are very unlikely to tell us anything about prudential value?
 - Hopefully, you’d doubt the truth of P1

4c. Eden Lin's expanded "experientially identical lifetime comparison argument"

1. If the vast majority of reasonable people have the brute intuition that A is higher in welfare than B, then it is reasonable to believe A is higher in welfare than B
2. The vast majority of reasonable people have the brute intuition that A is higher in welfare than B
3. Therefore, it is reasonable to believe A is higher in welfare than B (from 1 & 2)
4. If hedonism is true, then A is not higher in welfare than B
5. Therefore, it is reasonable to believe hedonism is false (from 3 & 4)

5. Background II: Psychology

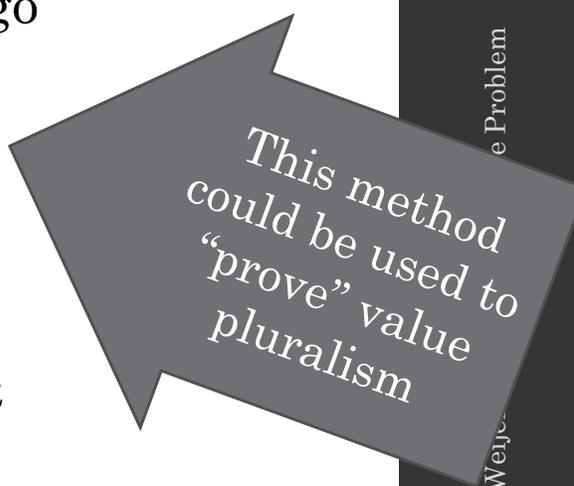
- Intuitions in psychology: gut feelings, valence but no labels
- Judgments: final view on something, *may* be faithfully expressed
- Bias: mental heuristic that inclines judgments to be inappropriate
- Biasing feature (of a thought experiment): a feature that often inclines judgments to be inappropriate, e.g. by triggering a mental heuristic to misfire, leading to an inclination toward an inappropriate judgment
- More on this available in Q&A!

6a. The “freebie” problem I

- Life 1 vs (identical) Life 2 + a freebie
- Some philosophers (e.g. Edin Lin): brute intuition that Life 2 is higher in value than Life 1 reveals that the freebie has greater value than zero
- Me: Every reasonable person should value Life 2 higher unless the freebie is not plausibly of value
 - Effectively equivalent in most cases: choose/prefer/brute-intuition-value-higher
- Reasonable people do not have 100% credence in their preferred theory of value
- Reasonable people have greater than 0% credence in other major theories of value
- ?Freebies of plausible value are viewed as valuable according to at least one major theory of value?

6b. The “freebie” problem II

- Basically, unless you are 100% (not rounded) sure that only experiences matter, you should prefer the non-deceived life to be on the safe side
- Compare:
- Theory: Only having your (restricted) desires satisfied makes life go well for you
- Test: 2 lives with exactly equal (restricted) desire satisfaction, but one has very high net affect and the other has very low net affect throughout life – which would you choose/should you prefer?
- Even a staunch supporter of the theory should choose the high net affect life
 - Failure to choose the happy life seems arrogant, irrational, or both



This method
could be used to
“prove” value
pluralism

6c. The “freebie” problem III

- This problem is common in ethics (and any place where values are compared)
- Solution:
- Set up comparisons more like the pre-Lin experience machine cases
- I.e., trade two rival goods against each other
- (Also need to discover and minimise the effect of biases like status quo bias) – I have a method for this
- How much more happiness is needed to avoid the freebie problem?
 - It depends on what credence in a theory of value is reasonable (95%?!)

7a. The Comparison argument suffers from the “freebie” problem

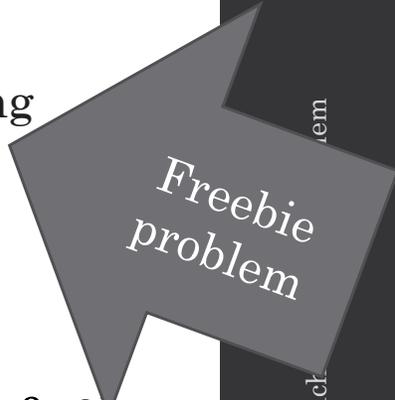
- Scenario: A and B have experientially identical lives, but B is in an experience machine (and has no people that care about or depend on him)
 1. A is higher in welfare than B
 2. If hedonism is true, then A is not higher in welfare than B.
 3. Therefore: Hedonism is false
- Life A: experiences 7/10, reality 9/10
- Life B: experiences 7/10, reality 1/10
- Chance of reality having positive intrinsic prudential value $> 0\%$ (not rounded)
- Therefore, Lin’s comparison argument has the “freebie” problem

7b. The Comparison argument is also verging on redundant

- There is already a famous thought experiment with the equal experience comparison set up:
- **The deceived businessman** (academic)
- Shelly Kagan, *Normative Ethics*, 1998, p34-36. (+ James Rachels etc.)
- Imagine 2 successful businessmen who died thinking that they had achieved everything they wanted: a loving wife, adoring children, a successful business, and the respect of the community. But 1 was completely wrong about his assessment of how things had gone: his wife was cheating, his children and the community at large were just using him for their own ends, and his business partner had been stealing from the business, which will soon be bankrupt.
- Which life would/should you prefer?
- Main difference is self-serving deception vs simulation

8a. Conclusion I

- It's doubtful that Lin's equal experience comparison argument enables us to learn anything about the value of pleasure or reality
 1. If the vast majority of reasonable people have the brute intuition that A is higher in welfare than B, then it is reasonable to believe A is higher in welfare than B
 - (unless there is good reason to think their beliefs are mistaken, e.g., by some biasing feature of the thought experiment)
 2. The vast majority of reasonable people have the brute intuition that A is higher in welfare than B
 3. Therefore, it is reasonable to believe A is higher in welfare than B (from 1 & 2)
 4. If hedonism is true, then A is not higher in welfare than B
 5. Therefore, it is reasonable to believe hedonism is false (from 3 & 4)



Freebie
problem

8b. Conclusion II

- **The best experience machine argument against hedonism should make a clear trade off between pleasure and reality without eliciting confounding biases... E.g....**
- Two people had lives very different to yours (reduce status quo bias)
- A lived his whole life in an experience machine and had net very positive experiences (the machine worked perfectly, he never knew)
- B lived his whole life in reality and had net slightly positive experiences
- Rate each life from 1-10 on how good it was for the person living it. (allows for equal ranking without encouraging “meh” answers)
- Unless the vast majority of acceptable* responses indicate A had the better life, it is reasonable to believe hedonism is false.

8c. Conclusion III

- **The “freebie” problem is a genuine problem and it crops up a lot in philosophy**
 - Lin’s comparison argument
 - The deceived businessman
 - Moore’s heap of filth
 - Turp’s argument: Turp, M. J. (2018). Normativity, Realism and Emotional Experience. *Philosophia*, 1-18.
 - And many more
- Look in your own area for powerful thought experiments that claim to isolate a value
- Please let me know of exmples!

Extra slides for question time

Background III: Judging scenarios

- Confabulation
 - We often use our deliberative cognition to justify our intuitions
 - The lawyer riding the elephant (Haidt)
 - It's not easy to tell when we are confabulating and when we are using “rationality” to come to a judgment
- What is “doing the work” in thought experiments/scenarios?
 - The features/values of philosophical interest
 - Irrelevant features (not related to the philosophical question)
 - How can we know?
- Are philosophers better than regular folk at judging scenarios?
 - Philosophers are experts at logical and coherent justifications
 - Do philosophers even have different judgments?

Background IV: Thought experiments and fitness-for-purpose

- Thought experiments
 - Research topic, e.g., “prudential well-being - what makes life go well?”
 - Research question, e.g., “are experiences all that matter for well-being?”
 - Hypothesis, e.g., “more than experiences matter for well-being”
 - Experiment (scenario), e.g., “imagine 2 lives... , which is better?”
 - Data (judgments), e.g., “the vast majority of people think that life 1 is better”
 - Conclusion, e.g., “the data show that more than experiences matter for well-being”
- Is a thought experiment-audience pairing fit for purpose?
 - Is the scenario (directly and only) relevant to the research question?
 - Can we trust the data (judgments about the scenario)?
 - How can we know?

Intuition to judgment (biases)

