

REVIEW OF *DEEP UTOPIA*

...the value of one's opinions, in a matter like this, is a function of how generously one has allowed the alternatives to play with one's soul.¹

People talk down the idea of living in a fool's paradise. But when one considers the nature of humanity, might it not seem that such a destination would be very suitable and desirable for us? I mean: If we are fools, then a fool's paradise would be exactly what we need.²

This is a meandering and playful book. I wasn't really in the right mindset to read it, but I enjoyed some parts and there was thought-provoking stuff throughout.

1. Space heaters

My favorite part was, easily, "The Exaltation of ThermoRex", a short story with the following premise:

Heißerhof, the country's leading industrialist, had bequeathed his vast fortune to a foundation established for the purpose of benefiting a particular portable electric room heater. We will refer to this room heater by its brand name, "ThermoRex". Heißerhof, who'd developed a reputation as being a bit of a misanthrope, had often been overheard saying that ThermoRex had done more for his welfare and comfort than any of his human companions ever had.³

2. Money

I like to think that increasing material prosperity will, because of the diminishing marginal value of money, eventually make people work less and enjoy more leisure time, as well as be more generous with others. Bostrom gives some reasons not to feel too sure of this, including:

Technological progress might create new ways of converting money into either quality or quantity of life, ways that don't have the same steeply diminishing returns that we experience today.

For example, suppose there were a series of progressively more expensive medical treatments that each added some interval of healthy life-expectancy, or that made somebody smarter or more physically attractive. For one million dollars, you can live five extra years in perfect health; triple that, and you can add a further five healthy years. Spend a bit more, and make yourself immune to cancer, or get an intelligence enhancement for yourself or one of your children, or improve your looks from a seven to a ten. Under these conditions—which could plausibly be brought about by technological advances—there could remain strong incentives to continue to work long hours, even at very high levels of income.⁴

3. Malthus

Obstacles to achieving utopia are *not* the main focus of the book, but it does spend some time on them. This includes some interesting discussion of population growth and the risk that it will—in the very long run—push society into a state where everyone lives at subsistence level. Bostrom thinks that, notwithstanding short-term trends for rich people to have fewer children, such a Malthusian condition will some day be inevitable unless it is averted by global population control.

Even space colonization can produce at best a polynomial growth in land, assuming we are limited by the speed of light—whereas population growth can easily be exponential, making this an ultimately unwinnable race. Eventually the mouths to feed will outnumber the loaves of bread to put in them, unless we exit the competitive regime of unrestricted reproduction. (Please note that this is a point about long-term dynamics, not a recommendation for what one country or another should be doing at present—which is an entirely different question altogether.)⁵

Robin Hanson tried to make subsistence-level existence sound not-terrible in *The Age of Em* ([review](#)), but I still find the idea pretty depressing. Bostrom gave me an extra reason to be depressed about it by explaining how technological advancement might (possibly) even reduce the level of welfare that a subsistence-level existence entails:

...you could have a model of fluctuating fortune within a life, where an individual dies if at any point their fortune dips below a certain threshold. In such a model, an individual may need to have a high *average* level of fortune in order to be able to survive long enough to successfully reproduce. Most times in life would thus be times of relative plenty.

In this model, inventions that smooth out fortune within a life—such as granaries that make it possible to save the surpluses when times are good and use them in times of need—lead to *lower* average well-being (while increasing the size of the population). This could be one of the factors that made the lives of early farmers worse than the lives of their hunter-gatherer forebears, despite the advance in technology that agriculture represented.⁶

4. Deep redundancy

The book's main focus is to consider what the implications would be if we someday reach the following state:

Technological maturity: A condition in which a set of capabilities exist that afford a level of control over nature that is close to the maximum that could be achieved in the fullness of time.⁷

This would include having superintelligent AGI and having extremely fine-grained control over the physical world, including our own bodies and minds.

What would we spend our time *doing* in such a world? Some people worry that no activity we could undertake would have any purpose any more. Bostrom distinguishes two versions of this:

The traditional and relatively superficial version of the purpose problem—let's call it *shallow redundancy*—is that human occupational labor may become obsolete due to progress in automation, which, with the right economic policies, would inaugurate an *age of abundance*. ...

The solution to shallow redundancy is to develop a leisure culture. Leisure culture would raise and educate people to thrive in unemployment. It would encourage rewarding interests and hobbies, and promote spirituality and the appreciation of the arts, literature, sports, nature, games, food, and conversation, and other domains...

The more fundamental version of the purpose problem...—let's call it *deep redundancy*—is that much leisure activity is *also* at risk of losing its purpose. ... It might even come to appear as though there would be no point in us doing anything—not working long hours for money, of course; but there would also be no point in putting effort into raising children, no point in going out shopping, no point in studying, no point in going to the gym or practicing the piano... et cetera.⁸

Why? Because technology could do it all better. Parenting: superintelligent parenting bots (presumably with human-like artificial bodies) could give children all the love and support they need, while being much better than we are at avoiding the infliction of any accidental psychological damage on the child. Shopping: AI models could know your preferences better than you do, and satisfy those preferences most effectively by making purchases without your involvement. Studying, gym, piano: machines and/or drugs could directly alter your mind and body to give you the skills without doing the work, and also to give you any desirable feelings that typically go along with doing the work.

Having recently read [a book on the Halting Problem and related topics](#), I'm primed to push back on the idea that a model could perfectly predict your preferences. *Maybe* that's possible—online advertising is sometimes pretty good at it already—but maybe we'll hit a point where further increases in accuracy aren't possible without basically running a full simulation of your brain, which might defeat the purpose.

Bostrom raises a related caveat when discussing brain editing:

In order to work out how to change the existing neural connectivity matrix to incorporate some new skill or knowledge, the superintelligent AI implementing the procedure might find it expedient to run simulations, to explore the consequences of different possible changes. Yet we may want the AI to steer clear of certain types of simulation because they would involve the generation of morally relevant mental entities, such as minds with preferences or conscious experiences. So the AI would have to devise the plan for exactly how to modify the subject's brain without resorting to proscribed types of computations. It is unclear how much difficulty this requirement adds to the task.⁹

Against the specter of deep redundancy, Bostrom proposes “[a] five-ringed defense”¹⁰ which I think I largely agree with. But the first ring, “hedonic valence”, is the one I'd put the most weight on: essentially, that even *if* we had no role in the future other than to passively enjoy the blessings prepared for us by our AI caretakers, that could in fact be a pretty amazing and wonderful future.

In addition to “purpose”, the book discusses related questions like whether we could find “fulfillment”, “richness”, and “meaning” in utopia. It's basically baked into the idea of “technological maturity” that we could *feel* like we had all those things: we could (more or less) make ourselves *feel* any way we wanted to by directly inducing the right states in our brains. But you might worry that something is lost if those feelings are not grounded in objective reality—if your life merely felt meaningful instead of being meaningful. Much of the book is devoted to seeking objective sources of fulfillment/meaning/etc that would not be undermined by the conditions of technological maturity. I respect the endeavor, but since I'm [drawn to a hedonistic theory of value](#) it's hard for me not to suspect that it's a bit of goose chase.

1. Nick Bostrom, *Deep utopia: life and meaning in a solved world* (Washington, DC: Ideapress Publishing, 2024), 5.
2. *Ibid.*, 299.
3. *Ibid.*, 366.
4. *Ibid.*, 9–10.
5. *Ibid.*, 56.
6. *Ibid.*, 25–26.
7. *Ibid.*, 62.
8. *Ibid.*, 147–49.
9. *Ibid.*, 140.
10. *Ibid.*, 151.

Posted 2024-09-07 by Jacob Williams on brokensandals.net. If you have feedback, [email me!](mailto:jacob@brokensandals.net)