The Case for (Strong) Longtermism

●●●

PHIL 1561 Ethics, Economics, and the Future
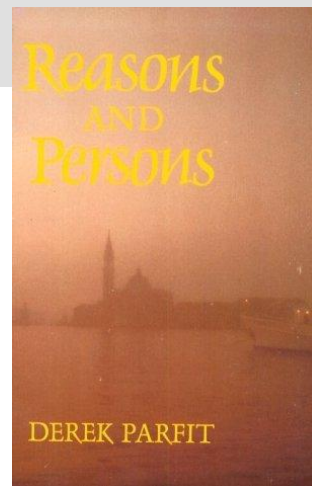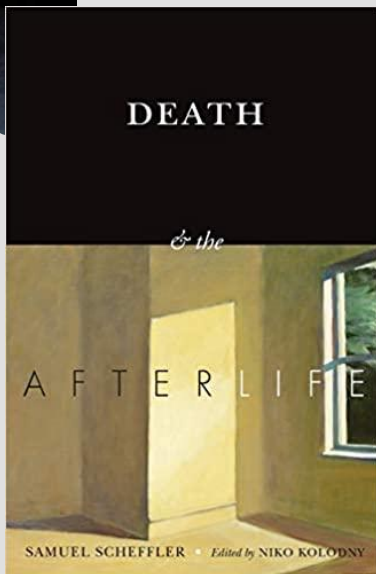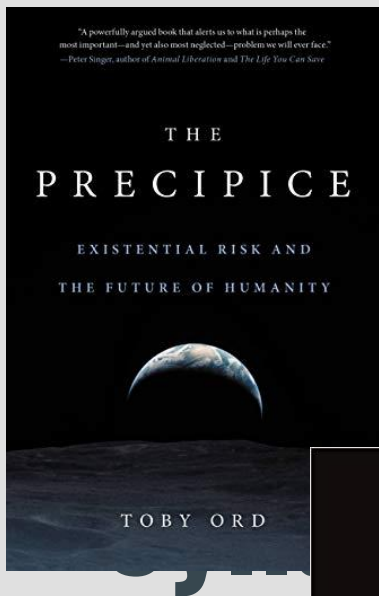Ryan Doody

Contents:
Syllabus Review
What is Longtermism?
The Case For It
The Case Against It

Course Syllabus
PHIL 1561 Ethics, Economics, and the Future
Spring 2024
9-10:20am

Instructor: Ryan Doody
Location: Sayles Hall 205
Office hours: Wednesday 1-2pm (and by appointment)
Course Website: https://canvas.brown.edu/courses/1094611
Email: ryan_doody@brown.edu

Teaching Assistant: Conrad Damstra (Conrad_Damstra@brown.edu)
Office hours: by appointment

Required Readings:
The Precipice: Existential Risk and the Future of Humanity, Toby Ord (Hachette Books, 2020) [$12.79]
Death and the Afterlife, Samuel Scheffler (Oxford University Press, 2013) [$21.47]
Other required readings will be posted on Canvas

Recommended (but not Required) Readings:
Reasons and Persons, Derek Parfit (Oxford University Press, 1984)
Why Worry About Future Generations?, Samuel Scheffler (Oxford University Press, 2018)
Intergenerational Justice, ed. Axel Gosseries and Lukas H. Meyer (Oxford University Press, 2009)

The Case for (Strong) Longtermism

The Case for (Strong) Longtermism

Impact on the **far future** is the most important feature of our actions today

What Is (Axiological Strong) Longtermism?

Axiological Strong Longtermism:
In the most important decision situations facing agents today,

(i)   Every option that is near-best overall is near-best for the far future.

(ii)   Every option that is near-best overall delivers much larger benefits in the far future than in the near future.

What Is (Axiological Strong) Longtermism?

Axiological Strong Longtermism:
In the most important decision situations facing agents today,

(i)   Every option that is near-best overall is near-best for the **far future**.

(ii) Every option that is near-best overall delivers much larger benefits in the **far future** than in the **near future**.

The Far Future?
Everything after some time t (where t is, e.g., 100 years after the point of decision).

The Near Future?
Everything before t and after the point of decision.

What Is (Axiological Strong) Longtermism?

Axiological Strong Longtermism:
In the most important decision situations facing agents today,

(i)   Every option that is **near-best overall** is **near-best for the far future**.

(ii)   Every option that is **near-best overall** delivers **much larger benefits** in the far future than in the near future.

Near-best overall / for the far future
Proportional distance from zero benefit to the maximal available benefit.

Much larger?
Multiplicative factor

Benefits?
Increases in value relative to the *status quo*

Why Think Longtermism is True?

Why Think It's True?

(Temporary) Assumptions:

1. *Ex Ante* Value of an option is its *expected value*.

2. Value is *total welfare*.

3. Time-separability for benefits. So, **V**(Overall benefits) = **V**(near-future benefits) + **V**(far-future benefits).

The *expected value* of an option is the weighted sum of the values of its outcomes, where the weights correspond to the probability that that outcome results.

Total Welfarism: The value of a complete world-history is the total sum of well-being in that world-history.

Roughly, *time-separability* means that the value of one period of time is independent of (or "separable" from) the values of other times.

Why Think It's True?

V(Near-future)

V(Far-future) = sum of each person's well-being

t

V(Overall) = **V**(Near-future) + **V**(Far-future)
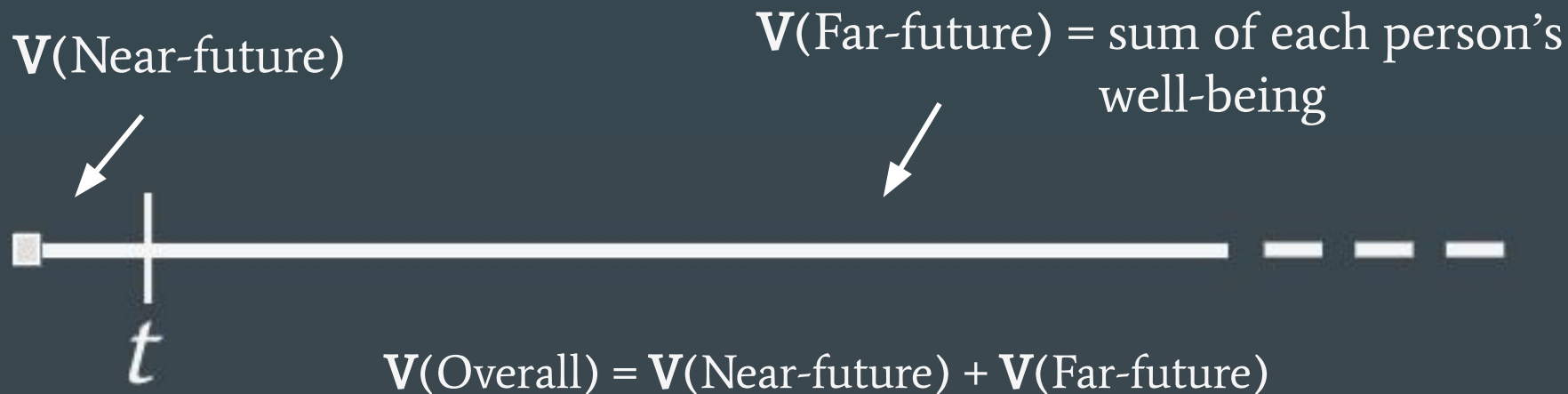
Why Think It's True?

There is (in expectation) a *vast* number of lives in the future of human civilization.

V(Near-future)

V(Far-future) = sum of each person's well-being

t

V(Overall) = **V**(Near-future) + **V**(Far-future)

There is (in expectation) a *vast* number of lives in the future of human civilization.

How vast?!

How vast the number of lives in the far-future?

10^4 centuries:
a million years

10^{10} lives per century:
ten billion lives

Scenario	Duration (centuries)	Carrying capacity (lives per century)	Number of future lives
Earth (mammalian reference class)	10^4	10^{10}	10^{14}
Earth (digital life)	10^4	10^{14}	10^{18}
Solar System	10^8	10^{19}	10^{27}
Solar System (digital life)	10^7	10^{23}	10^{30}
Milky Way	10^{11}	10^{25}	10^{36}
Milky Way (digital life)	10^{11}	10^{34}	10^{45}

How vast the number of lives in the far-future?

Space Settlement (Solar System)

"If humanity lives not only on Earth but also on other planets—in our own solar system, elsewhere in the Milky Way, or in other galaxies too—then terrestrial constraints on future population size disappear, and astronomically larger populations become possible."

"Even if we only settle **the solar system**, civilisation would have **over 5 billion years** until the end of the main sequence lifetime of the Sun, and we would have access to over two billion times as much sunlight power as we do now."

Scenario	Duration (centuries)	Carrying capacity (lives per century)	Number of future lives
Earth (mammalian reference class)	10^4	10^{10}	10^{14}
Earth (digital life)	10^4	10^{14}	10^{18}
Solar System	10^8	10^{19}	10^{27}
Solar System (digital life)	10^7	10^{23}	10^{30}
Milky Way	10^{11}	10^{25}	10^{36}
Milky Way (digital life)	10^{11}	10^{34}	10^{45}

How vast the number of lives in the far-future?

Space Settlement (Milky Way)

"If humanity lives not only on Earth but also on other planets—in our own solar system, elsewhere in the Milky Way, or in other galaxies too—then terrestrial constraints on future population size disappear, and astronomically larger populations become possible."

"If we are able to widely settle the rest of **the Milky Way**, then we could access well over **250 million** rocky habitable-zone planets, each of which has the potential to support **trillions of lives** over the course of their sun's lifetimes."

Scenario	Duration (centuries)	Carrying capacity (lives per century)	Number of future lives
Earth (mammalian reference class)	10^4	10^{10}	10^{14}
Earth (digital life)	10^4	10^{14}	10^{18}
Solar System	10^8	10^{19}	10^{27}
Solar System (digital life)	10^7	10^{23}	10^{30}
Milky Way	10^{11}	10^{25}	10^{36}
Milky Way (digital life)	10^{11}	10^{34}	10^{45}

How vast the number of lives in the far-future?

Digital Sentience

"The second radical possibility is that of digital sentience: that is, conscious artificial intelligence (AI)."

"[I]t makes interstellar travel much easier: it is easier to sustain digital than biological beings during very long-distance space travel. [And] digital sentience could dramatically increase the number of beings who could live around one star: digital agents could live in a much wider variety of environments, and could more efficiently turn energy into conscious life."

Scenario	Duration (centuries)	Carrying capacity (lives per century)	Number of future lives
Earth (mammalian reference class)	10^4	10^{10}	10^{14}
Earth (digital life)	10^4	10^{14}	10^{18}
Solar System	10^8	10^{19}	10^{27}
Solar System (digital life)	10^7	10^{23}	10^{30}
Milky Way	10^{11}	10^{25}	10^{36}
Milky Way (digital life)	10^{11}	10^{34}	10^{45}

How vast the number of lives in the far-future?

Scenario	Duration (centuries)	Carrying capacity (lives per century)	Number of future lives
Earth (mammalian reference class)	10^4	10^{10}	10^{14}
Earth (digital life)	10^4	10^{14}	10^{18}
Solar System	10^8	10^{19}	10^{27}
Solar System (digital life)	10^7	10^{23}	10^{30}
Milky Way	10^{11}	10^{25}	10^{36}
Milky Way (digital life)	10^{11}	10^{34}	10^{45}

Estimate: 10^{24}

(Low estimate: 10^{18})

Objection 1: The Washing-out Hypothesis

The Washing-out Hypothesis

The far-future effects of one's actions are very hard to predict. So, in expectation, the effects of one's decision on the near-future carry more weight than the effects on the far-future.

"Might it be that the expected instantaneous value differences between available actions **decay with time** from the point of action, and **decay sufficiently fast** that in fact the near-future effects tend to be the most important contributor to expected value?"

The Washing-out Hypothesis

The far-future effects of one's actions are very hard to predict. So, in expectation, the effects of one's decision on the near-future carry more weight than the effects on the far-future.
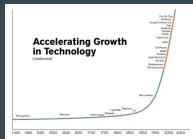
Response:
There are things we can do now that we can be fairly confident will affect the far-future in positive ways.

Namely,

(1) **mitigating the risks of premature human extinction**, and

(2) **positively shaping the development of artificial superintelligence**

Existential Risk

What influence do we have?

The Washing-out Hypothesis

The far-future effects of one's actions are very hard to predict. So, in expectation, the effects of one's decision on the near-future carry more weight than the effects on the far-future.
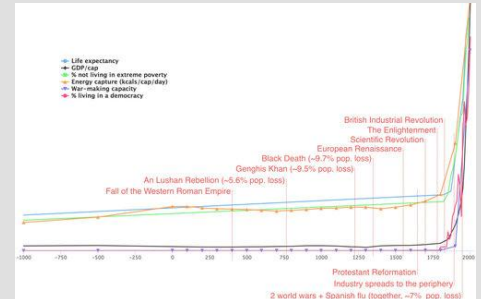
The Hinge of History

"We live during **the hinge of history**. Given the scientific and technological discoveries of the last two centuries, the world has never changed as fast. We shall soon have even greater powers to transform, not only our surroundings, but ourselves and our successors. If we act wisely in the next few centuries, humanity will survive its most dangerous and decisive period. Our descendants could, if necessary, go elsewhere, spreading through this galaxy."

Derek Parfit, *On What Matters, Vol. II* (2011)

Discussion Question:
Is Parfit right that "we live during the hinge of history"?

Discussion Que...

Is Parfit right th... during the hinge... history"?

The 100-Year Extinction Panic Is Back, Right on Schedule

🎁 Share free access   ⤷   🔖   💬 853

By Tyler Austin Harper

Mr. Harper is an assistant professor of environmental studies at Bates College.

Jan. 26, 2024

"Do you think we'll need to buy guns?" The student's question seemed to drop the temperature in the room by several degrees. I was at a dinner with fellow academics, a few college students and a guest speaker who had just delivered an inspiring talk about climate justice.

Objection 2: The argument rests on many controversial assumptions

Controversial Assumptions

(Temporary) Assumptions:

1. *Ex Ante* Value of an option is its *expected value.*

2. Value is *total welfare.*

3. Time-separability for benefits. So, **V**(Overall benefits) = **V**(near-future benefits) + **V**(far-future benefits).

What about risk-aversion?

Making people happy vs. making happy people?

Assuming a 0% rate of pure time preference.

Objection 3: Epistemic worries

Cluelessness

"Perhaps the beings that are around will be very unlike humans. Perhaps their societies, if they have anything that can be called a society at all, will be organized in enormously different ways. For these and other reasons, perhaps the kinds of things that are conducive to the well-being of far-future creatures are very different from the kinds of things that are conducive to our well-being. Given all of this, can we really have **any clue** about the far-future value of our actions *even in expectation?*"

"[W]e are *clueless* both about what the far future will be like, and about the differences that we might be able to make to that future."

We will discuss these more later on.

Deontic Strong Longtermism:
One ought to choose the option that's best for the very far future.

The Stakes Sensitivity Argument

P1 If the stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor, one ought to choose a near-best option.

P2 In the most important decisions facing agents today, the stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor.

C In the most important decisions facing agents today, one ought to choose a near-best option.

Consequentialism:
One ought to do what's best.

Deontology:
in some cases, we aren't required to do what's best (we have the **prerogative** not to); and, in some cases, we shouldn't do what's best (e.g., because it violates a "**side-constraint**").

The Stakes Sensitivity Argument

P1 If the stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor, one ought to choose a near-best option.

P2 In the most important decisions facing agents today, the stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor.

_____

C In the most important decisions facing agents today, one ought to choose a near-best option.

Discussion Question:
Suppose you have a rich friend who has left their wallet unattended. You could easily swipe a few hundred dollars—they're so rich they probably won't even notice—and donate it to your favorite Longtermist cause.

Should you?

The Stakes Sensitivity Argument

P1 If the stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor, one ought to choose a near-best option.

P2 In the most important decisions facing agents today, the stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor.

C In the most important decisions facing agents today, one ought to choose a near-best option.

Discussion Question:

Suppose you have a rich friend who has left their *crypto* wallet unattended. You could easily swipe a few hundred *bitcoin* ~~dollars~~—they're so rich they probably won't even notice—and donate it to your favorite Longtermist cause.

Should you?

What Do You Think about Longtermism?

Next Time:
How valuable is existential risk reduction?