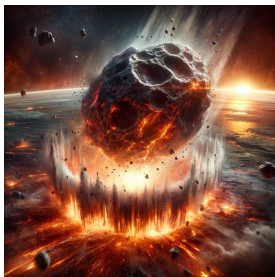


The Optimistic Meta-Induction

Petra Kosonen



Should we lower our credences in near-term existential risks because of past failed doomsday predictions?

tinyurl.com/optimistic-induction

Existential risks

- ▶ Some argue that existential risk mitigation should be one of the most important global priorities.¹
- ▶ Existential risks are risks that threaten the destruction of humanity's long-term potential.
- ▶ Such risks might be posed by, for example, synthetic pathogens, artificial general intelligence, climate change or asteroids.
- ▶ Extinction risks are one type of existential risk.

Existential risks

Risks that threaten the destruction of humanity's long-term potential.

¹See for example Bostrom (2003), Bostrom (2013) and Greaves and MacAskill (2021).

Existential risk mitigation

- ▶ The justification for existential risk mitigation need not rely on a high credence in these risks.²
- ▶ As humanity's future is potentially very long—and thus might contain vast amounts of value—even small decreases in the net probability of existential catastrophe may correspond to enormous increases in expected moral value.³
- ▶ However, many do, in fact, give relatively high probabilities for these risks materializing.

²In fact, Thorstad (2023) argues a high credence in these risks undermines the case for existential risk mitigation.

³Bostrom (2013).

Existential risk estimates

- ▶ Oxford philosopher Toby Ord estimates that the probability of an existential catastrophe occurring by 2120 is 1 in 6.⁴
- ▶ British Astronomer Royal Sir Martin Rees is even more pessimistic, estimating only a 50% chance that present civilization on Earth will survive until the end of this century.⁵
- ▶ The Doomsday Clock now stands at 90 seconds to midnight—closest to global catastrophe it has ever been.⁶
- ▶ Nearly one-third (31%) of Americans and 23% of Brits believe an apocalyptic disaster is very to somewhat likely during their lifetime.⁷

⁴Ord (2020, p. 167).

⁵Rees (2003, p. 8).

⁶See Mecklin (2023).

⁷YouGov (2015a, p. 1) and YouGov (2015b, p. 10).

The Optimistic Meta-Induction

- ▶ However, our estimates of the probabilities of existential risks might be undermined by the *Optimistic Meta-Induction*:⁸

The Optimistic Meta-Induction

The history of humanity is full of doomsday predictions that turned out to be wrong, so we have no reason to believe that our current doomsday predictions are approximately right.

- ▶ In this talk, I'll explore whether we should lower our subjective credences in near-term existential risks in light of the historical track record of failed predictions.
- ▶ Should we turn back the Doomsday Clock?

⁸The Optimistic Meta-Induction is analogous to the *Pessimistic Meta-Induction* in philosophy of science: Given that past widely accepted scientific theories were found to be false, we have no reason to think our current scientific theories are true or approximately true.

Failed doomsday predictions

- ▶ Here are some examples of predicted doom:
- ▶ Very many Christians have attempted to predict the end of time and the second coming of Jesus: Martin Luther, Christopher Columbus, John Napier (mathematician), Jakob Bernoulli (mathematician)...⁹
- ▶ In fact, stories about the end of the world are prevalent across cultures: the tale of Noah and the flood, the Norse myth of Ragnarök, the Hindu myths of world annihilation and regeneration, the Zoroastrian, Babylonian, Sumerian, Buddhist, Islamic, Greek, Roman, African, Mayan and Native American myths describing the destruction and transformation of the world.¹⁰

⁹See e.g., Weber (1999) and Aveni (2016).

¹⁰Wojcik (1997, p. 5).

One World or None

- ▶ More recently, many prominent people predicted a nuclear armageddon:
- ▶ **Albert Einstein (1952):** “Only the creation of a world government can prevent the impending self-destruction of mankind.”¹¹
- ▶ **Bertrand Russell:**¹² “Before the end of the present century, unless something quite unforeseeable occurs, one of three possibilities will have been realized. These three are: —
 1. The end of human life, perhaps of all life on our planet.
 2. A reversion to barbarism after a catastrophic diminution of the population of the globe.
 3. A unification of the world under a single government, possessing a monopoly of all the major weapons of war.”
- ▶ **Scientist and novelist C. P. Snow (1961):** “Within, at the most, ten years, some of those [nuclear] bombs are going off. I am saying this as responsibly as I can. *That* is the certainty.”¹³

¹¹Nathan and Norden (1960, p. 566).

¹²Russell (1951).

¹³Weaver et al. (1961, p. 259).

Population Bomb

- ▶ Paul and Anne Ehrlich famously predicted in their 1968 book, “The Population Bomb,” that by the 1970s the world would see widespread famines and hundreds of millions of deaths due to overpopulation.¹⁴
- ▶ In 1970, Paul Ehrlich wrote a letter from an imagined future in which nearly four billion lives had been lost over a fifteen year period¹⁵ and sixty-five million Americans had been starved to death in the 1980s,¹⁶ leading to a 1999 population of the “United States of North America” of 22.6 million.¹⁷
- ▶ In a 1969 talk, he stated: “If I were a gambler, I would take even money that England will not exist in the year 2000.”¹⁸

¹⁴Ehrlich (1968).

¹⁵Ehrlich (1970, p. 25).

¹⁶Ehrlich (1970, pp. 23–24).

¹⁷Ehrlich (1970, p. 23).

¹⁸Dixon (1971).

Earth Day 1970 - SKIP

- ▶ Ecologist Kenneth E. F. Watt (University of California at Davis) told Time in 1970 that “[a]t the present rate of nitrogen buildup, it’s only a matter of time before light will be filtered out of the atmosphere and none of our land will be usable.”¹⁹
- ▶ He continued that “California’s air pollution is already so bad that it may start a wave of mass deaths by 1975.”
- ▶ Harvard University Nobel laureate biologist George Wald warned in 1970 that “[c]ivilization will end within 15 or 30 years unless immediate action is taken against problems facing mankind.”²⁰

¹⁹Shavitz (1970, p. 59).

²⁰Williams (2015, p. 374).

Doomed to fail: Were past doomsday predictions wrong?

- ▶ One might object to the Optimistic Meta-Induction by arguing that past doomsday predictions were, in fact, not wrong.
- ▶ There is one obvious way in which past doomsday predictions were wrong: the predicted catastrophes never happened.

Evidence 1

The predicted catastrophes never happened.

Probabilistic predictions

- ▶ One might object that the doomsday predictions were probabilistic: rather than predicting certainty of doom, they gave it a high probability.
- ▶ And sometimes one might give an event a high subjective probability and not be epistemically wrong (in expectation) even if the event does not happen—sometimes the unlikely happens.
- ▶ So, the fact that human extinction has never happened does not mean that past predictions were epistemically unjustified.
- ▶ However, one can argue that the predictions must have been wrong because it would be very unlikely for humanity to still exist had these predictions been right about the magnitude of the risk.

Evidence 2

It would be very unlikely for humanity to still exist had these predictions been right about the magnitude of the risk.

Observation selection effects

- ▶ But, given that we cannot observe doomsday predictions ever materializing, judgement about their wrongness is subject to observation selection effects.²¹
- ▶ Had they been right in the past, there would be no one to observe it.
- ▶ It could be that extinction risk was high and most civilizations destroy themselves relatively quickly, but we are the lucky survivors.
- ▶ Observation selection effects explain why it is not surprising that humanity exists, even if extinction risk was high.
- ▶ So, we cannot use the continued existence of humanity as evidence against past doomsday predictions.

²¹Ćirković et al. (2010). For a contrary view, see Thomas (2024).

Lack of near misses

- ▶ However, we can change the argument slightly: had they been right about the magnitude of the risk in the past, we would expect to at least find evidence of near misses—but we do not.
- ▶ Observation selection effects do not explain the lack of near misses.

Evidence 3

It would be very unlikely for humanity to still exist and *not have experienced any near misses* had these predictions been right about the magnitude of the risk.

Nuclear near misses

- ▶ But, contrary to Evidence 3, humanity might have experienced near misses.
- ▶ There is disagreement about how close we came to a nuclear war during the Cold War, but some scholars have argued that we were very lucky in avoiding one.
- ▶ John F. Kennedy himself estimated that the Cuban missile crisis had somewhere between one-in-three and even chance of leading to a nuclear war.²² But ExComm member McGeorge Bundy estimated that probability to be 1%.²³
- ▶ Lundgren (2013) estimates that the expected probability of nuclear war over the past sixty-six years was greater than 50%.
- ▶ However, Mueller (2010, 2014) and Waltz (Sagan and Waltz, 1995) argue that the risks have been exaggerated and the probability of a nuclear war has always been very low.

²²See Blight and Welch (1989, p. 84).

²³Blanton (1997).

Past predictions were unscientific

- ▶ Lastly, we have other evidence for past doomsday predictions having been wrong: given what we now know, they could never have happened (or had a tiny probability of happening).
- ▶ For example, many doomsday predictions were religious stories that conflict with the scientific worldview.
- ▶ We do not need to know that human extinction has not happened to know that these kind of doomsday predictions were wrong.
- ▶ For example, we can imagine that the predicted events are supposed to happen in our future—we would still know that they are wrong.

Evidence 4

Given what we now believe, the predicted catastrophes could never have happened (or had a tiny probability of happening).

Does the wrongness of past doomsday predictions mean current doomsday predictions are wrong?

- ▶ We have four kinds of evidence against past doomsday predictions: the predicted catastrophes never happened, it would be unlikely for humanity to still exist had they been right about the magnitude of the risk, the lack of near misses and, given what we now believe, the predicted catastrophes could never have happened (or had a tiny probability of happening).
- ▶ One might accept that past doomsday predictions were wrong, but insist that this does not mean our current doomsday predictions are wrong.
- ▶ Next, I will discuss objections to the Optimistic Meta-Induction of this sort.

Past predictions not scientifically informed

- ▶ A plausible objection to the Optimistic Meta-Induction is that people who made doomsday predictions in the past did not study the risks rigorously and instead relied on, for example, religious stories.²⁴
- ▶ Existential risks have only recently—in the post-atomic age—become the object of rigorous, quantitative, and scientifically serious study.²⁵
- ▶ Therefore, the argument goes, past doomsday predictions are not the right reference class for reasoning about our current existential risk estimates.
- ▶ Even if they were wrong in the past, that does not mean we are wrong now; nowadays we know better.

²⁴Moynihan (2020, pp. 20–21, 27) points this out.

²⁵Moynihan (2020, pp. 20, 27), Bostrom (2013, p. 27) and Ord (2020, p. 62).

Doomsday bias

- ▶ But the prevalence of past doomsday predictions might be taken to show that we have a tendency to exaggerate existential risks:²⁶

Doomsday bias

Some (or most) people have a tendency to make doomsday predictions, whether or not they are justified.

- ▶ The same psychological tendencies that led people to create apocalyptic stories in the past could also influence today's scientists and philosophers, leading them to overestimate the risks we face.
- ▶ But a relatively low credence in near-term existential risks is plausibly less affected by the Doomsday Bias.

²⁶This tendency might be tight to specific cultures too.

Past predictions not about extinction

- ▶ The next objection argues that past doomsday predictions are not the right reference class for reasoning about current doomsday predictions because past predictions concerned global catastrophes or apocalyptic events rather than human extinction.
- ▶ Moynihan (2020, p. 32): “Haven’t humans been predicting the end of time since the beginning of history? Certainly—but extinction has nothing to do with religious apocalypse.”
- ▶ However, some past doomsday predictions did focus on the possibility of extinction rather than global catastrophes or religious apocalypses: e.g., nuclear war and environmental degradation.
- ▶ Also, the same psychological tendencies may predispose people to believe in both apocalypses and extinction because apocalypses and human extinction share similar features.

Expand reference class to include global catastrophes

- ▶ Another objection to the Optimistic Meta-Induction suggests that we should expand the reference class to include global catastrophes.
- ▶ When we do so, it is no longer true that all past doomsday predictions failed: many global catastrophes have indeed happened.
- ▶ However, my claim is not that global catastrophes or extinction are impossible, but rather that some of us may have a tendency to overestimate their probabilities.
- ▶ So, even if some past doomsday predictions were correct (when we understand doomsday predictions to include global catastrophes as well), it is still plausibly the case that most predicted global catastrophes did not happen.

Cyclic view - SKIP

- ▶ Many cultures have believed that humanity would persist indefinitely, as they held a circular view of time.
- ▶ Does this challenge the Optimistic Meta-Induction? It seems not.
- ▶ Some cultures might have a tendency to produce doomsday predictions that other cultures lack. So we might still be vulnerable to the Doomsday Bias.
- ▶ I am not defending the statement 'people of all cultures and times have always made doomsday predictions.'
- ▶ Also, cultures with a cyclical view of time often still predicted global catastrophes, which share similar features with existential risks.
- ▶ So, the same psychological tendencies might be at work again.

- ▶ The next objection to the Optimistic Meta-Induction is that, although there is a history of failed doomsday predictions, we should not conclude that current doomsday predictions are wrong because nowadays we can ask talented forecasters who have a good track record to estimate the risks.
- ▶ And, if these 'superforecasters' say there is a high probability of an existential catastrophe, we should believe them.
- ▶ If existential risks are similar to the shorter-run geopolitical forecasting questions studied in previous research, superforecasters could be a more accurate guide to what will actually happen.²⁷
- ▶ What do the superforecasters say about existential risks?

²⁷Karger et al. (2023, p. 14).

Forecasting

- ▶ The median superforecaster predicted a 9% chance of global catastrophe (that kills at least 10%) and a 1% chance of extinction by year 2100.²⁸
- ▶ The median superforecaster also gave the following estimates for extinction risks from various causes by year 2100:
 1. AI extinction: 0.38%
 2. Engineered pathogen extinction: 0.01%
 3. Nuclear extinction: 0.074%
 4. Total extinction risk: 1%
- ▶ There was significant internal disagreement among superforecasters.
- ▶ However, if anything, superforecasters seem to have relatively low credences for near-term extinction. So, superforecasters do not support a high credence in human extinction in this century.

²⁸Karger et al. (2023, p. 4).

True past doomsday predictions impossible

- ▶ Another objection to the Optimistic Meta-Induction is that we could never find accurate doomsday predictions in the past: the situation is such that we can only find failed ones.
- ▶ So, of course, we should not be surprised to find only failed predictions in the past.
- ▶ However, we could find no past doomsday predictions at all.
- ▶ Alternatively, we might only find doomsday predictions made by people who are psychologically very different to oneself.
- ▶ If one has evidence of not sharing the same psychological characteristics, then—possibly—the Optimistic Meta-Induction would not apply to oneself.

Preventing catastrophes - SKIP

- ▶ Another objection is that the reason why past predictions proved wrong was that people took steps to prevent the catastrophes.
- ▶ For some potential extinction events, such as a global nuclear war, substantial effort has been invested into preventing them.
- ▶ But, similarly as in the past, we may act to prevent the predicted catastrophes.
- ▶ Those announcing that doom is near may be subject to *Sleepwalk Bias*: The tendency to postulate that people will sleepwalk into a disaster.²⁹
- ▶ Additionally, for many doomsday predictions, no action was taken—or even could have been taken—to prevent them, because they were never truly possible in the first place.

²⁹Schubert (2016).

Arguments for high existential risk

- ▶ Next, I will discuss some reasons to think that near-term existential risk is high.

Specific evidence

- ▶ The Optimistic Meta-Induction is higher-order evidence, that is, evidence about the character of first-order evidence.
- ▶ But it seems somewhat arrogant to claim that the carefully formulated existential risk estimates are wrong based solely on a meta-argument when the scientists and philosophers who made those estimates have studied the topics carefully.
- ▶ For example, Toby Ord provides detailed descriptions of where the 1/6 existential risk estimate comes from.
- ▶ The Optimistic Meta-Induction is insensitive to evidence about particular risks.
- ▶ However, higher-order evidence can be informative, even though first-order evidence is also important.
- ▶ This type of higher-order evidence can be especially valuable when our first-order evidence is uncertain and ambiguous, as is the case with many existential risks.

Species have gone extinct - SKIP

- ▶ Another objection to the Optimistic Meta-Induction is that although humans have not gone extinct, other species have.
- ▶ However, the target of the Optimistic Meta-Induction is a high subjective credence in existential risks in the near term.
- ▶ And, based on estimates of extinction risk from natural causes, humanity's expected future lifespan is at least 87,000 years.³⁰
- ▶ On the other hand, the average lifespan of hominin species is approximately one million years.
- ▶ So, evidence from natural history does not support a high credence in near-term human extinction.

³⁰Snyder-Beattie et al. (2019).

Civilizations have gone extinct

- ▶ One objection to the Optimistic Meta-Induction is that although humans have not gone extinct, many civilizations have.
- ▶ And, as Paul Ehrlich said in an interview: “Civilisations have collapsed before: the question is whether we can avoid the first time [an] entire global civilisation has given us the opportunity of having the whole mess collapse.”³¹
- ▶ How long have civilizations lasted in the past?

³¹Ehrlich (2011).

How long have civilizations lasted in the past?

- ▶ Even the longest lasting empires seem to be relatively short-lived (up to 7 centuries), and the average lifespan of a civilization is 336 years.³²
- ▶ But some people often survived their civilization's collapse, so we should not draw conclusions about extinction risk from these numbers.
- ▶ Furthermore, one could argue that pre-Industrial Revolution civilizations are irrelevant in this context, as society has fundamentally changed—we are in a different reference class.

³²Kemp (2019).

Doomsday Argument - SKIP

- ▶ Another reason to think that near-term existential risk is high is the Doomsday Argument.
- ▶ First consider the *Self-Sampling Assumption*:

The Self-Sampling Assumption

You should reason as if you were a random sample from the set of all observers (in your reference class).

- ▶ Next suppose we have two hypotheses:³³
 1. Doom Early: humanity goes extinct sometime within this century and the total number of humans that will have lived is 200 billion.
 2. Doom Late: humanity survives this century and eventually settles the galaxy. The total number of humans who will ever have lived is 200 trillion.

³³The argument presented here is from Bostrom (2008).

Doomsday Argument - SKIP

- ▶ Now you discover that your birth rank is around 60 billion—that is roughly the number of humans who have lived before you.
- ▶ This gives you reason to consider the “Doom Early” scenario more probable than you previously believed: it would be surprising to find ourselves so early in humanity’s history if the “Doom Late” scenario were correct.
- ▶ Therefore, you have reason to expect that human extinction happens relatively soon.
- ▶ I will not evaluate the Doomsday Argument in this talk; however, there is considerable disagreement about whether it should be accepted.
- ▶ Also, the conclusion of the Doomsday Argument is that the future of humanity is relatively short. But it might still be some hundreds of years—too distant in the future to validate our shorter-term existential risk predictions.

The Vulnerable World Hypothesis

- ▶ A thought experiment that concerns the potential risks associated with technological advancement.³⁴

The Vulnerable World Hypothesis

If technological development continues, it may eventually reach a set of capabilities that make the destruction of civilization highly probable, unless extremely strict preventive policing is implemented.³⁵

- ▶ The idea is that as technology advances, the ability of individual actors to cause harm increases, making society more vulnerable.

³⁴Bostrom (2019). See also von Neumann (1955).

³⁵Bostrom (2019, p. 457).

Urn of invention - SKIP

- ▶ The central idea illustrated using a metaphorical 'urn of invention'.³⁶
- ▶ Each new idea, discovery and technology humanity produces is like drawing a ball from this urn.
- ▶ Most balls are white (representing beneficial technologies), some are gray (technologies with mixed consequences), but there is a concern that, someday, we might draw a "black ball": a technology so dangerously powerful and easily misused that it could result in our destruction.

³⁶Bostrom (2019, p. 455).

Urn of invention - SKIP

- ▶ Furthermore, the reason we have not drawn a black ball is not due to exceptional caution on our part; we have simply been lucky.
- ▶ We can imagine an alternative history where nuclear weapons are easily produced using readily available materials.³⁷
- ▶ Social order would probably be severely disrupted by draconian measures to prevent dangerous individuals from obtaining these materials or societies might descend into nuclear civil wars.

³⁷Bostrom (2019, pp. 456–457).

The Vulnerable World Hypothesis

- ▶ A reason for cautious optimism is that, as technology advances and becomes more powerful, a larger group of people is typically needed to weaponize it for hostile purposes.³⁸
- ▶ The more people needed to weaponize a technology, the more effective societal controls become at mitigating, reducing or preventing potential harm.
- ▶ Consequently, it is unlikely that a lone individual or a small group could cause human extinction.
- ▶ The Vulnerable World Hypothesis may not be true, but it would be unreasonable to be confident that it is false.³⁹
- ▶ I find the Vulnerable World Hypothesis the most plausible argument for relatively high near-term existential risk.

³⁸Kevin Kelly, in personal communication with Steven Pinker. See Pinker (2019, p. 302). See also Kelly (2012).

³⁹Bostrom (2019, p. 458).

Objective chances

- ▶ Lastly, I will discuss some lessons from the Optimistic Meta-Induction.
- ▶ It seems plausible that we should not penalize our credences in existential risks when we know their objective chances.
- ▶ For example, we can estimate the frequency of asteroid collisions using historical data, so we have less reason to believe that Doomsday Bias influences our estimates in this case.⁴⁰

⁴⁰Ćirković et al. (2010) argue that using historical data to estimate extinction risks tends to underestimate their probabilities due to the presence of observation selection effects.

Mixed counterargument

- ▶ I find the most convincing argument against the Optimistic Meta-Induction to be a mixed one: either past doomsday predictions belong to a different reference class than current ones (as they were not based on careful thinking), or observation selection effects explain humanity's continued existence.
- ▶ So, it is possible to look at the history of failed doomsday predictions without making significant updates based on it.
- ▶ However, the many failed religious doomsday predictions might still reveal something about our psychology—specifically, that we have a bias toward exaggerating existential risks.

Existential risks possible

- ▶ In discussing the Optimistic Meta-Induction, it is important to remember that existential risks are, of course, possible.
- ▶ How should we reason when our first-order evidence could be influenced by the Doomsday Bias?
- ▶ As Pinker (2019, p. 294) writes on the Y2K: “The Great Y2K Panic does not mean that all warnings of potential catastrophes are false alarms, but it reminds us that we are vulnerable to techno-apocalyptic delusions.”
- ▶ My conclusion is modest: while existential catastrophes are possible, we should be careful before concluding that near-term existential risk is high.
- ▶ There is probably something to the Optimistic Meta-Induction.

Mitigating existential risks - SKIP

- ▶ As I mentioned in the beginning of this talk, one need not have a high subjective credence in existential risks in order to act to mitigate them.
- ▶ Given the potential size of the future, much is at stake.⁴¹
- ▶ So, even small decreases in overall existential risk could lead to enormous gains in expected moral value.⁴²
- ▶ And, as Moynihan (2020, p. 82) writes: “We now know that there are only very small slivers of our galactic surroundings that are compatible with supporting life; we all recognise that life is a cosmic rarity, a fragile oasis of activity in a universe that is otherwise mostly extinct matter and void—and this is why we are inclined to take X-risk deadly seriously.”

⁴¹Bostrom (2013).

⁴²Bostrom (2013). For a contrary view, see Thorstad (2023).

Conclusion

- ▶ Many people assign relatively high probabilities to near-term existential risks.
- ▶ But our estimates of existential risk probabilities might be undermined by the Optimistic Meta-Induction: history of humanity is full of doomsday predictions that turned out to be wrong, so we have no reason to believe that our current predictions are approximately right.
- ▶ I discussed various objections to the Optimistic Meta-Induction, such as observation selection effects explaining the continued existence of humanity and past doomsday predictions being the wrong reference class to reason about current doomsday predictions.
- ▶ While existential risks are possible, all in all I conclude that we should be careful before assigning a high subjective credence to near-term existential risks.

Arguments discussed

- ▶ Past doomsday predictions were not wrong:
 1. The doomsday predictions were probabilistic.
 2. Observation selection effects.
 3. We had near misses (with e.g., nuclear weapons).
- ▶ Even if they were wrong in the past, it does not follow we are wrong:
 1. Past predictions not scientifically informed.
 2. Past predictions not about extinction.
 3. Expand reference class to include global catastrophes.
 4. Cyclic view.
 5. Forecasting.
 6. True past doomsday predictions impossible.
 7. Preventing catastrophes.
- ▶ Evidence for high near-term existential risk:
 1. There is evidence for particular risks.
 2. In the past civilizations and species have gone extinct.
 3. The Doomsday Argument.
 4. The Vulnerable World Hypothesis.

References I

- Aveni, A. (2016), *Apocalyptic Anxiety: Religion, Science, and America's Obsession with the End of the World*, University Press of Colorado, Boulder, Colorado.
- Blanton, T. (1997), 'Annals of blinksmanship', https://nsarchive2.gwu.edu/nsa/cuba_mis_cri/annals.htm. The Wilson Quarterly.
- Blight, J. G. and Welch, D. A. (1989), *On the Brink: Americans and Soviets Reexamine the Cuban Missile Crisis*, Hill and Wang, New York.
- Bostrom, N. (2003), 'Astronomical waste: The opportunity cost of delayed technological development', *Utilitas* **15**(3), 308–314.
- Bostrom, N. (2008), 'The doomsday argument', *Think* **6**(17–18), 23–28.
- Bostrom, N. (2013), 'Existential risk prevention as global priority', *Global Policy* **4**(1), 15–31.
- Bostrom, N. (2019), 'The vulnerable world hypothesis', *Global Policy* **10**(4), 455–476.

References II

Ćirković, M. M., Sandberg, A. and Bostrom, N. (2010), 'Anthropic shadow: Observation selection effects and human extinction risks', *Risk Analysis* **30**(10), 1495–1506.

Dixon, B. (1971), 'In praise of prophets', *New Scientist* p. 606.

URL: <https://archive.org/details/screenshot-of-ehrlichs-claim-that-england-will-not-exist>

Ehrlich, P. (2011), 'Paul ehrlich, a prophet of global population doom who is gloomier than ever'.

URL: <https://www.theguardian.com/environment/2011/oct/23/paul-ehrllich-global-collapse-warning>

Ehrlich, P. R. (1968), *The Population Bomb*, Ballantine Books, New York.

Ehrlich, P. R. (1970), 'Looking backward from 2000 a.d.', *The Progressive* **34**(4), 23–25.

References III

- Greaves, H. and MacAskill, W. (2021), 'The case for strong longtermism'. Global Priorities Institute Working Paper No. 5–2021.
URL: <https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/>
- Karger, E., Rosenberg, J., Jacobs, Z., Hickman, M., Hadshar, R., Gamin, K., Smith, T., Williams, B., McCaslin, T. and Tetlock, P. E. (2023), 'Forecasting existential risks: Evidence from a long-run forecasting tournament'. Forecasting Research Institute Working Paper No. 1.
- Kelly, K. (2012), 'Myth of the lone villain'.
URL: <https://kk.org/thetechnium/myth-of-the-lon/>
- Kemp, L. (2019), 'The lifespans of ancient civilisations'.
URL: <https://www.bbc.com/future/article/20190218-the-lifespans-of-ancient-civilisations-compared>
- Lundgren, C. (2013), 'What are the odds?: Assessing the probability of a nuclear war', *The Nonproliferation Review* **20**(2), 361–374.

References IV

Mecklin, J. (2023), 'A time of unprecedented danger: It is 90 seconds to midnight'.

URL: <https://thebulletin.org/doomsday-clock/current-time/>

Moynihan, T. (2020), *X-Risk: How Humanity Discovered Its Own Extinction*, Urbanomic, Falmouth.

Mueller, J. (2010), *Atomic Obsession: Nuclear Alarmism from Hiroshima to Al-Qaeda*, Oxford University Press,, New York.

Mueller, J. (2014), 'Fire, fire. (review of e. schlosser's 'command and control')', Times Literary Supplement.

Nathan, O. and Norden, H. (1960), *Einstein on Peace*, Simon and Schuster, New York.

Ord, T. (2020), *The Precipice: Existential Risk and the Future of Humanity*, Bloomsbury, London.

Pinker, S. (2019), *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*, Penguin Books, New York.

References V

- Rees, M. (2003), *Our Final Hour: A Scientist's Warning: How Terror, Error, and Environmental Disaster Threaten Humankind's Future in This Century—on Earth and Beyond*, Basic Books, New York.
- Russell, B. (1951), 'The future of man', *Atlantic*.
- Sagan, S. D. and Waltz, K. N. (1995), *The Spread of Nuclear Weapons: A Debate*, W. W. Norton & Company, New York.
- Schubert, S. (2016), 'Sleepwalk bias, self-defeating predictions and existential risk'.
URL:
<https://www.lesswrong.com/posts/gEShPto3F2aDdT3RY/sleepwalk-bias-self-defeating-predictions-and-existential>
- Shavitz, B. (1970), 'Fighting to save the earth from man'.
URL: <https://time.com/vault/issue/1970-02-02/page/60/>
- Snyder-Beattie, A., Ord, T. and Bonsall, M. (2019), 'An upper bound for the background rate of human extinction', *Scientific Reports* **9**(1), 11054.

References VI

- Thomas, T. (2024), 'Dispelling the anthropic shadow'. Global Priorities Institute Working Paper No. 20–2024.
URL: <https://globalprioritiesinstitute.org/dispelling-the-anthropic-shadow-teruji-thomas/>
- Thorstad, D. (2023), 'High risk, low reward: A challenge to the astronomical value of existential risk mitigation', *Philosophy and Public Affairs* **51**(4), 373–412.
- von Neumann, J. (1955), 'Can we survive technology?', *Fortune* pp. 106–152.
- Weaver, W., Snow, C. P., Hesburg, T. M. and Baker, W. O. (1961), 'The moral un-neutrality of science', *Science* **133**(3448), 255–262.
- Weber, E. (1999), *Apocalypses: Prophecies, Cults, and Millennial Beliefs through the Ages*, Harvard University Press, Cambridge, Massachusetts.
- Williams, W. E. (2015), *American Contempt for Liberty*, Hoover Institution Press, Stanford, California.

References VII

Wojcik, D. (1997), *The End of the World As We Know It: Faith, Fatalism, and Apocalypse in America*, New York University Press, New York.

YouGov (2015a), 'Apocalyptic disaster', https://ygo-assets-websites-editorial-emea.yougov.net/documents/InternalResults_150309_apocalyptic_disaster_Website.pdf.

YouGov (2015b), 'Disaster 02/27/15', https://ygo-assets-websites-editorial-emea.yougov.net/documents/tabs_OPI_disaster_20150227_QaIqpU3.pdf.