

# *The Optimistic Meta-Induction\**

Petra Kosonen<sup>†</sup>

September, 2025

ABSTRACT: The justification for existential risk mitigation need not rely on a high subjective credence in these risks. However, many do, in fact, give relatively high probabilities for these risks materializing in the near future. In this paper, I argue that our estimates of existential risk probabilities might be undermined by 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 our current doomsday predictions are approximately right. I explore whether we should lower our subjective credences in near-term existential risks in light of the historical track record of failed predictions. I discuss 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. I conclude that while existential risks and global catastrophes are of course possible, we may have a tendency to exaggerate such risks.

---

\*For valuable feedback and discussions, I wish to thank Tomi Francis, Dmitri Gallow, Johan Gustafsson, Stefan Schubert, Christian Tarsney, Philip Trammell, and the audiences of 13th Oxford Workshop on Global Priorities Research, the 17th International Congress on Logic, Methodology and Philosophy of Science and Technology, and the PPE Society's Eighth Annual Meeting.

<sup>†</sup>Population Wellbeing Initiative, University of Texas at Austin. I would be grateful for comments: kosonenpetra@gmail.com

Some argue that existential risk mitigation should be one of the most important global priorities.<sup>1</sup> Existential risks are risks that endanger humanity’s long-term potential. These risks could be posed by, for example, synthetic pathogens, artificial general intelligence, climate change or asteroids. Extinction risks are one type of existential risk. The justification for existential risk mitigation need not rely on a high credence in these risks.<sup>2</sup> As humanity’s future could be 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.<sup>3</sup>

However, many do, in fact, give relatively high probabilities for these risks materializing. Oxford philosopher Toby Ord estimates that the probability of an existential catastrophe occurring within the next 100 years is 1 in 6.<sup>4</sup> 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.<sup>5</sup> And the Doomsday Clock now stands at 90 seconds to midnight—the closest to global catastrophe it has ever been.<sup>6</sup>

This message has also been heard by the public, as a quarter (24%) of those surveyed in the United States, Canada, Australia and the United Kingdom believe that the probability of human extinction within a century is 50% or greater.<sup>7</sup> And

---

<sup>1</sup>See for example Bostrom (2003), Bostrom (2013) and Greaves and MacAskill (2021).

<sup>2</sup>Thorstad (2023) argues that a high subjective credence in existential risks is unlikely to justify the overwhelming importance of mitigating them unless it is combined with the time of perils hypothesis.

<sup>3</sup>Bostrom (2013).

<sup>4</sup>Ord (2020, p. 167).

<sup>5</sup>Rees (2003, p. 8).

<sup>6</sup>The Doomsday Clock is operated by the Bulletin of the Atomic Scientists, founded in 1945 by Albert Einstein and University of Chicago scientists who helped develop the first atomic weapons in the Manhattan Project. The Doomsday Clock was created two years later, and it is set every year in consultation with Nobel laureates. See Mecklin (2023). Pinker (2019, p. 309), a critic of the Doomsday Clock, writes: “The Doomsday Clock, despite adorning a journal with “Scientists” in its title, does not track objective indicators of nuclear security; rather, it’s a propaganda stunt intended, in the words of its founder, “to preserve civilization by scaring men into rationality.”

<sup>7</sup>Randle and Eckersley (2015).

a YouGov poll found that nearly one-third (31%) of Americans and 23% of Brits believe an apocalyptic disaster is very to somewhat likely during their lifetime.<sup>8</sup> A more recent poll found that 34% of Americans believe that the end of the human race on Earth will occur in the next 100 years.<sup>9</sup>

In this paper I argue that our estimates of the probabilities of existential risks might be undermined by the *Optimistic Meta-Induction*:<sup>10</sup>

**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.

This paper explores 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?

The paper proceeds as follows: §1 explores historical doomsday predictions. §2 examines whether past doomsday predictions were wrong. §3 discusses whether the wrongness of past doomsday predictions implies that current doomsday predictions are also wrong. §4 goes through arguments for a high net existential risk. §5 considers lessons we should take from the Optimistic Meta-Induction. §6 concludes.

---

<sup>8</sup>YouGov (2015a, p. 1) and YouGov (2015b, p. 10).

<sup>9</sup>YouGov (2023).

<sup>10</sup>The Optimistic Meta-Induction is analogous to the *Pessimistic Meta-Induction* in philosophy of science: Given that past widely accepted scientific theories turned out to be false, we have no reason to think our current scientific theories are true or approximately true. See Laudan (1981) and Psillos (2022) on the Pessimistic Meta-Induction.

# 1 The boy who cried apocalypse: History of doomsday predictions

History is full of failed doomsday predictions: the 2012 Mayan apocalypse prediction<sup>11</sup>, Y2K<sup>12</sup>, new ice age, aliens, the killer bees, collision of the Earth with a comet<sup>13</sup>, and so on.<sup>14</sup> This section presents some examples of apocalyptic predictions.

## 1.1 Historical examples of doomsday predictions

“When the thousand years are completed, Satan will be released from his prison (Revelation 20:7)”, says an ominous prophecy in the Book of Revelation.<sup>15</sup> Many Christians have attempted to predict the end of time and the second coming of Jesus.<sup>16</sup> The list includes some famous individuals as well. Martin Luther, who started the Protestant Reformation, believed the end time was near, no later than 1600.<sup>17</sup> Christopher Columbus, influenced by the writings of Joachim of Fiore,<sup>18</sup> predicted that the world would end in 1656 once all humans would have been converted to Christianity.<sup>19</sup> John Napier, a mathematician, calculated that the Last Judgment would be in 1688 or 1700.<sup>20</sup> According to Jakob Bernoulli, a mathe-

---

<sup>11</sup>Aveni (2016, pp. 3–6).

<sup>12</sup>There were widespread predictions of a computer bug that would crash many computers in the first seconds of year 2000 and lead to major catastrophes worldwide. See, for example, Hughes (2008, pp. 82–83).

<sup>13</sup>Aveni (2016, p. 7).

<sup>14</sup>See Ebell and Milloy (2019) for a list of failed doomsday predictions.

<sup>15</sup>Revelation (2017, p. 1102).

<sup>16</sup>See, for example, Weber (1999), Hall (2009) and Aveni (2016, p. 60).

<sup>17</sup>Weber (1999, p. 66).

<sup>18</sup>Aveni (2016, p. 69).

<sup>19</sup>McGovern (1992, p. 17) and McIver (1999, p. 32).

<sup>20</sup>Weber (1999, p. 92) and McIver (1999, p. 36).

matician, a comet would destroy the Earth in 1719.<sup>21</sup>

Like Christians, early Muslims also expected the Last Judgment to come soon, seventy or a hundred years after Muhammad and his followers emigrated from Mecca in 622.<sup>22</sup> 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, Greek, Roman, African, Mayan and Native American myths describing the destruction and transformation of the world.<sup>23</sup>

## 1.2 Recent examples of doomsday predictions

ONE WORLD OR NONE.<sup>24</sup> “Only the creation of a world government can prevent the impending self-destruction of mankind”, predicted Albert Einstein in 1952.<sup>25</sup> J. Robert Oppenheimer echoed similar spirit when he proclaimed that, in the long run, “without world government, there could be no permanent peace, and without peace there would be atomic warfare.”<sup>26</sup> Bertrand Russell, in turn, predicted the following: “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

---

<sup>21</sup>McIver (1999, p. 46).

<sup>22</sup>Hall (2009, pp. 35–41).

<sup>23</sup>Wojcik (1997, p. 5).

<sup>24</sup>This was the slogan of the movement that advocated for a global government to prevent a nuclear war. See Mueller (2010, p. 74).

<sup>25</sup>Nathan and Norden (1960, p. 566).

<sup>26</sup>Bird and Sherwin (2005, pp. 341–342).

of all the major weapons of war.”<sup>27</sup> In 1961, the scientist and novelist C. P. Snow declared: “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.”<sup>28</sup> In 1976, the computer scientist Joseph Weizenbaum said: “I am completely certain—there is not the slightest doubt in my mind—that by the year 2000, you [students] will all be dead.”<sup>29</sup> Asked about the future of international relations, the political scientist Hans Morgenthau responded in 1979: “Future, what future? I am extremely pessimistic. In my opinion the world is moving ineluctably towards a third world war—a strategic nuclear war. I do not believe that anything can be done to prevent it.”<sup>30</sup>

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.<sup>31</sup> They argued that the growth in global population was unsustainable and would outstrip food production and resources. In 1970, Paul Ehrlich wrote a letter from an imagined future in which nearly four billion lives had been lost over a fifteen-year period<sup>32</sup> and sixty-five million Americans had been starved to death in the 1980s,<sup>33</sup> leading to a 1999 population of the “United States of North America” of 22.6 million.<sup>34</sup> In a 1969 talk given at the Institute of Biology in London, he even went so far as to state: “If I were a gambler, I would take even money that England will not exist in the year 2000.”<sup>35</sup>

---

<sup>27</sup>Russell (1951).

<sup>28</sup>Weaver et al. (1961, p. 259).

<sup>29</sup>Cited in Pinker (2019, p. 309) as an address to the incoming graduate students, Faculty of Arts and Sciences, Harvard University, September 1976.

<sup>30</sup>Boyle (1985, p. 73). In the same discussion, Morgenthau concedes that he was in a very pessimistic mood that day and might give a different answer another day.

<sup>31</sup>Ehrlich (1968). See also Malthus (1798)

<sup>32</sup>Ehrlich (1970, p. 25).

<sup>33</sup>Ehrlich (1970, pp. 23–24).

<sup>34</sup>Ehrlich (1970, p. 23).

<sup>35</sup>Dixon (1971)

EARTH DAY 1970. 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.”<sup>36</sup> He continued that “California’s air pollution is already so bad that it may start a wave of mass deaths by 1975.” And 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.”<sup>37</sup>

To summarize, several doomsday predictions have been made throughout history. On the Earth Day 1970 predictions, Bailey (2000) writes that “[t]he prophets of doom were not simply wrong, but spectacularly wrong.” The next section explores whether or not past doomsday predictions were wrong.

## 2 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. This section presents four arguments for why past doomsday predictions were wrong.

There is one obvious way in which past doomsday predictions were wrong: the predicted catastrophes never happened. As a matter of fact, things turned out such that humanity still exists, contrary to predictions.

**Evidence 1:** The predicted catastrophes never happened.

However, the doomer could object that the doomsday predictions were probabilistic: rather than predicting certainty of doom, they assigned 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—

---

<sup>36</sup>Shavitz (1970, p. 59).

<sup>37</sup>Williams (2015, p. 374).

sometimes the unlikely happens. So, the fact that human extinction has never happened does not mean that past predictions were epistemically unjustified.

The optimist can respond 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 the past doomsday predictions been right about the magnitude of the risk.

There is, however, a way for the doomer to respond: Given that we cannot observe doomsday predictions ever materializing, judgment about their wrongness is subject to observation selection effects.<sup>38</sup> 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.<sup>39</sup>

Now the optimist 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 the past doomsday predictions been right about the magnitude of the risk.

---

<sup>38</sup>Carter (1974), Bostrom (2002), Tegmark and Bostrom (2005), Ćirković (2008) and Ćirković et al. (2010).

<sup>39</sup>For a contrary view, see Thomas (2024). Thomas (2024, p. 3) argues that “the existence of the anthropic shadow is essentially irrelevant to estimating risks.” Thomas’s argument is that the fact that life has continued so long is evidence that the rate of extinction events is low, and that this cancels out the observation selection effect. Consequently, the true rate of potential extinction events is close to the observed frequency.

However, contrary to Evidence 3, humanity may 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 lucky in avoiding one. Lundgren (2013) estimates that the expected probability of a nuclear war over the past sixty-six years (before the publication of the book) 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.<sup>40</sup>

One dangerous incident happened during the Cuban missile crisis in 1962.<sup>41</sup> On October 27th, the U.S. Navy tried to force a Soviet submarine to surface by deploying practice depth charges. Unbeknownst to the Americans, the submarine was armed with a nuclear torpedo with an explosive yield comparable to that of the Hiroshima bomb.<sup>42</sup> The submarine had not surfaced for two days, and the latest news they had received led them to believe that hostilities between the U.S. and the Soviet Union could break out at any moment.<sup>43</sup>

“It felt like you were sitting in a metal barrel, which somebody is constantly blasting with a sledgehammer”, remembers Vadim Orlov, a communications intelligence officer of the submarine.<sup>44</sup> Inside the submarine, only the emergency lights were operational, the temperature ranged from 45 to 60 degrees Celsius, CO2 levels had risen to dangerous levels, and crew members started fainting “like dominoes.”<sup>45</sup> This went on for four hours, until the Americans seemed to strike the submarine with something more powerful.<sup>46</sup> This enraged the exhausted Second

---

<sup>40</sup>For lists of nuclear incidents and close calls, see Union of Concerned Scientists (2015), Conn (2016) and Baum et al. (2018). See also Sagan (1993). On the probability of a nuclear war, see Intriligator and Brito (1981), Bereanu (1983), Avenhaus et al. (1989), Brito and Intriligator (1996), Hellman (2008), Barrett et al. (2013), Borrie (2014), Lewis et al. (2014), Baum (2015) and Barrett (2016).

<sup>41</sup>Mozgovoi (2002). See also Huchthausen (2002).

<sup>42</sup>Burr and Blanton (2002).

<sup>43</sup>Savranskaya (2005).

<sup>44</sup>Mozgovoi (2002).

<sup>45</sup>Mozgovoi (2002).

<sup>46</sup>Mozgovoi (2002).

Captain Valentin Savitsky, who ordered the crew to prepare the nuclear torpedo for battle readiness.<sup>47</sup> Savitsky shouted: “We’re going to blast them now! We will die, but we will sink them all—we will not disgrace our Navy!”<sup>48</sup> However, Second Captain Vasili Arkhipov vetoed the decision to launch the nuclear torpedo against the will of the other two officers on board.<sup>49</sup>

This description of events is controversial.<sup>50</sup> But, according to Rear Admiral Carl J. Seiberlich, the U.S. would have made a “nuclear counter-response” had the Soviets used nuclear torpedoes.<sup>51,52</sup> 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.<sup>53</sup> Yet some disagree: ExComm member McGeorge Bundy estimated that probability to be just 1%.<sup>54</sup>

Let’s now turn to the final type of evidence against past doomsday predictions. With many of these predictions, we do not need to know that human extinction has not happened to know that they were wrong. For example, we can imagine that the predicted events are supposed to happen in our future.<sup>55</sup> We would still know that they are wrong because we have other evidence for them being wrong: given what we now know, they could never have happened (or had a tiny probability of

---

<sup>47</sup>Mozgovoi (2002).

<sup>48</sup>Mozgovoi (2002).

<sup>49</sup>Burr and Blanton (2002).

<sup>50</sup>Burr and Blanton (2002).

<sup>51</sup>Burr and Blanton (2002, note 9).

<sup>52</sup>Another famous incident happened on 26th September 1983, when a Soviet satellite issued a warning of an incoming missile launched from the U.S. towards the Soviet Union. Contrary to his orders, the officer on duty, Stanislav Petrov, chose not to report the matter further suspecting it was a false alarm. His instincts proved right, as the satellite had mistakenly detected sunlight reflected off clouds. See Hoffman (2011, pp. 6–11). More recently, according to Watergate journalist Bob Woodward, Donald Trump said that “you don’t know how close we were to war” with North Korea in 2017. See Walker (2020).

<sup>53</sup>See Blight and Welch (1989, p. 84).

<sup>54</sup>Blanton (1997). ExComm (Executive Committee of the National Security Council) was composed of United States government officials that advised President John F. Kennedy during the Cuban Missile Crisis in 1962.

<sup>55</sup>Thank you Tomi Francis for this case.

happening). For example, many doomsday predictions were religious stories that conflict with the scientific worldview.

**Evidence 4:** Given what we now believe, the predicted catastrophes could never have happened (or had a tiny probability of happening).

To summarize, 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).

### **3 Does the wrongness of past doomsday predictions imply current ones are wrong too?**

One might accept that past doomsday predictions were wrong, but insist that this does not mean that our current existential risk estimates are wrong. This section discusses objections to the Optimistic Meta-Induction of this sort.

#### **Unscientific theories**

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.<sup>56</sup> Existential risks have only recently—in the post-atomic age—become the object of rigorous, quantitative, and scientifically serious study.<sup>57</sup> 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.

---

<sup>56</sup>Moynihan (2020, pp. 20–21, 27) also points this out.

<sup>57</sup>Moynihan (2020, pp. 20, 27), Bostrom (2013, p. 27) and Ord (2020, p. 62).

However, many scientists and philosophers have also made failed doomsday predictions: Albert Einstein, Bertrand Russell, C. P. Snow, and so on. Therefore, even scientifically minded people have sometimes exaggerated the risks.

Moreover, future generations may regard our current existential risk estimates as ‘unscientific’ by their standards.<sup>58</sup> We can imagine a future in which the scientists and philosophers of each generation assign a high credence to existential risks, but each generation also has more advanced science and technology. With improved science and technology, each future generation will be in a better position to estimate existential risks. Nevertheless, a future scientist or philosopher who holds a high credence in some existential risk should revise their confidence given the history of failed doomsday predictions. One should not ignore the Optimistic Meta-Induction just because each generation has more advanced science. However, the doomer may respond that only the doomsday predictions made since the start of the scientific revolution, or since science matured sufficiently, should count. And, the argument continues, there have not been that many doomsday predictions since that happened.

But 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.<sup>59</sup> The prevalence of past doomsday predictions might be taken to show that we have a tendency to exaggerate existential risks:<sup>60,61</sup>

**Doomsday Bias:** A significant number of people have a tendency to make doomsday predictions, whether or not they are justified.<sup>62</sup>

---

<sup>58</sup>Thank you Phil Trammell for this point.

<sup>59</sup>A relatively low credence in near-term existential risks is plausibly less affected by Doomsday Bias.

<sup>60</sup>This tendency might be tied to specific cultures too. Gross and Gilles (2012, p. 12) write: “The widespread belief in some kind of apocalypse is the defining cultural phenomenon of our time, and its popularity has something profound to tell us about the American mind-set...”

<sup>61</sup>On how cognitive biases might cause us to *underestimate* existential risks, see Yudkowsky (2008).

<sup>62</sup>Note that, on the Doomsday Bias, it is not necessarily the case that most people have a ten-

One piece of evidence for this tendency is that belief in a doomsday scenario is common, yet there is disagreement about what will cause the end. As Gross and Gilles (2012, p. 9) observe: “In America, everyone believes in the apocalypse. The only question is whether Jesus or global warming will get here first.” Also, when experts and forecasters are asked to estimate existential risks, predictions are highly correlated across topics. For example, those who are more concerned about artificial intelligence are also more concerned about pandemics and nuclear war.<sup>63</sup> This suggests that some people are more prone than others to believe in doomsday scenarios.

### **Past predictions not about extinction**

The next objection to the Optimistic Meta-Induction is 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; there have been few truly apocalyptic movements or religions that predict an absolute end of history.<sup>64</sup>

As Moynihan (2020, p. 32) observes, “Haven’t humans been predicting the end of time since the beginning of history? Certainly—but extinction has nothing to do with religious apocalypse.” Moynihan (2020, p. 32) also argues that the ‘end of the world’ is not the same as ‘human extinction,’ explaining that “[e]xtinction means that the wider universe continues without our species”, whereas “this is generally not a feature of apocalyptic narratives...”

However, some past doomsday predictions did focus on the possibility of extinction rather than global catastrophes or religious apocalypses. For example, predictions regarding nuclear war, overpopulation, and environmental degradation concerned human extinction. Also, the same psychological tendencies may pre-

---

endency to believe in doomsday predictions. If one finds oneself having a high credence in existential risks, one might have the psychological predisposition to believe in them (which most people lack).

<sup>63</sup>Karger et al. (2023, p. 4).

<sup>64</sup>Hughes (2008, p. 78).

dispose 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 relevant reference class of doomsday predictions to include all global catastrophes. And when we do so, it is no longer true that all past doomsday predictions failed: many global catastrophes have indeed happened.<sup>65</sup>

However, the optimist can change the argument slightly: our current predictions of global catastrophes are *probably* wrong because most historical predictions of global catastrophes were wrong.

**The Optimistic Meta-Induction\*:** Most past predictions of global catastrophes turned out to be wrong, so most of our current predictions of global catastrophes are probably wrong too.

Even if some past doomsday predictions were correct (when we understand doomsday predictions to include global catastrophes as well), it is still the case that most predicted global catastrophes did not happen. Yet this may not comfort the doomer. Sure, each individual prediction is probably wrong, they might concede. But even a single correct one may mean the end of humanity—so we should be worried. The optimist can agree and still insist that the risks have been exaggerated. The claim is not that global catastrophes or extinction are impossible, but rather that some of us may have a tendency to overestimate their probabilities—perhaps especially so when it comes to extinction.

---

<sup>65</sup>Thank you Dmitri Gallow for suggesting this kind of objection.

## Cyclic view

In many cultures people believed that humanity would persist indefinitely, as they held a circular view of time. Does this challenge the Optimistic Meta-Induction?<sup>66</sup> It seems not.

Some cultures may have a tendency to produce doomsday predictions that other cultures lack. So, *we* might be vulnerable to the Doomsday Bias even if those with a cyclical view of time are not. Also, cultures with a cyclical view of time still predicted global catastrophes, which share similar features with existential risks. So, the same psychological tendencies to exaggerate global catastrophes might influence them too.

## Forecasting

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. Superforecasters could be a more accurate guide to what will actually happen if existential risks are similar to the shorter-run geopolitical forecasting questions studied in previous research.<sup>67</sup> What do the superforecasters say about existential risks?

The median superforecaster predicted a 9% chance of global catastrophe (that kills at least 10%) and a 1% chance of extinction by year 2100.<sup>68</sup> The median superforecaster also gave the following estimates for extinction risks from various causes

---

<sup>66</sup>Thank you Johan Gustafsson for this objection.

<sup>67</sup>Karger et al. (2023, p. 14).

<sup>68</sup>Karger et al. (2023, p. 4). Karger et al. (2023, p. 4, n.4) define a catastrophic event as one causing the death of at least 10% of humans alive at the beginning of a five-year period and extinction as reduction of the global population to less than 5000. The mean total extinction risk estimate of superforecasters was 3.73%. See Karger et al. (2023, p. 21).

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.<sup>69</sup>

## **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.<sup>70</sup> So—of course—we should not be surprised to only find failed predictions in the past.

However, we could find no past doomsday predictions at all. So, the set up does not make it necessary that we find failed doomsday predictions in the past. Alternatively, we might only find doomsday predictions made by people who are psychologically very different to us. If one has evidence of not sharing the same psychological characteristics as those who made the failed predictions, then the Optimistic Meta-Induction would possibly not apply to oneself. Furthermore, the prevalence of doomsday predictions could be taken as evidence that a significant number of people have a tendency to make doomsday predictions.

---

<sup>69</sup>It is possible that forecasting existential risks demands a different set of skills than those used for predicting short-run topics. For example, forecasters often rely on base rates, which might be unavailable in the case of long-run risks. See Karger et al. (2023, p. 14).

<sup>70</sup>Thank you Johan Gustafsson for this objection.

## Preventing catastrophes

Someone might grant that history is filled with doomsday predictions that turned out wrong and that our current existential risk predictions belong to the same reference class as the past predictions. Yet they might argue that we should not sigh in relief: the reason why those predictions turned out wrong was that people took steps to prevent the catastrophes.

Substantial effort has been invested into preventing some global catastrophes, such as a nuclear war. Many analysts believe that potential nuclear catastrophes were avoided thanks to proactive measures taken by governments, corporations and individuals.<sup>71</sup> Similarly, in response to the Y2K threat, governments and companies invested massive amounts of capital in upgrades to their computer networks to prevent the catastrophe from happening.<sup>72</sup> Yet another example is the global response to the discovery of the hole in the ozone layer.<sup>73</sup>

However, just like in the past, we may also 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.<sup>74</sup> Secondly, no useful action was taken, or even could have been taken, to prevent many of the predicted catastrophes. The reason is that these catastrophes were never truly possible in the first place.

To summarize, I have discussed various arguments that challenge the Optimistic Meta-Induction. Many of these arguments posit that the Optimistic Meta-Induction uses the wrong reference class because, for example, past predictions were not scientific; or they were not about extinction; or we should expand the reference class to include global catastrophes; or past theories were cyclic; or nowadays we can ask forecasters to estimate existential risks. I also discussed the arguments suggesting that we could never find accurate doomsday predictions in the

---

<sup>71</sup>Hughes (2008, p. 83).

<sup>72</sup>Gross and Gilles (2012, p. 24).

<sup>73</sup>Gross and Gilles (2012, p. 26)

<sup>74</sup>Schubert (2016).

past and that past doomsday predictions turned out wrong due to preventive actions.

## 4 Arguments for high existential risk

In this section, I will discuss some reasons to think that near-term existential risk is high. The Optimistic Meta-Induction asks us to be careful before assigning high subjective credences to near-term existential risks. But what if evidence for high risks is convincing? I will discuss a few reasons to think that existential risk is high: evidence for particular risks, in the past civilizations and species have gone extinct, the Doomsday Argument and the Vulnerable World Hypothesis.

### Specific evidence

The Optimistic Meta-Induction is higher-order evidence, that is, evidence about the character of first-order evidence. It may seem overconfident 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.<sup>75</sup>

However, this type of higher-order evidence can be especially valuable when our first-order evidence is uncertain and ambiguous, as is often the case with many existential risks.<sup>76</sup> There seems to be a dubious correlation between the subjectivity of the evidence and how high the risk estimate is. For example, Ord (2020, p. 167) gives the following estimates for existential catastrophes from various causes by year 2120: 1 in 1,000,000 from asteroid or comet impact, 1 in 30 from engi-

---

<sup>75</sup>Analogously, some argue that the strongest arguments for scientific realism are developed by focusing on first-order evidence derived directly from scientific investigation. See Chakravarty (2017, §1.3).

<sup>76</sup>Thank you Stefan Schubert for this point.

neered pandemics and 1 in 10 from unaligned artificial intelligence. It seems that the greater the risk is estimated to be, the more subjective the evidence for it is. This might be explained by the Doomsday Bias: unless rigorous evidence suggesting a low risk is available, we tend to overestimate the risks. The greatest risks are anthropogenic and relatively recent, which makes it more challenging to obtain evidence that is not highly subjective—so there is room for exaggeration. However, anthropogenic risks might also truly be more dangerous than non-anthropogenic ones.

### **Species have gone extinct**

Another objection to the Optimistic Meta-Induction is that although humans have not gone extinct, other species have. But, based on estimates of extinction risk from natural causes, humanity’s expected future lifespan is at least 87,000 years.<sup>77</sup> 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. Note that the target of the Optimistic Meta-Induction is a high subjective credence in existential risks in the *near term*. After all, eventually humanity will probably go extinct.

### **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.”<sup>78</sup> We now have one global civilization, so we should be worried.

How long have civilizations lasted in the past? The longest lasting empires from 600 B.C. to 600 A.D. were the Parthian-Sassanid empire (lasted 7 centuries) fol-

---

<sup>77</sup>Snyder-Beattie et al. (2019).

<sup>78</sup>Ehrlich (2011).

lowed by Rome, Andhra, the Byzantine and Ptolemaic empires (lasted around 3 to 4 centuries).<sup>79</sup> And the longest lasting empires from 3000 to 600 B.C. were Egypt's New and Old Empires (lasted 5 centuries each) and Hsia-Shang empire (lasted 4 centuries).<sup>80</sup> Even the longest lasting empires seem to be relatively short-lived. And the average lifespan of a civilization is 336 years.<sup>81</sup> However, some individuals often survived the collapse of their civilization, so we should not draw conclusions about extinction risk from these numbers. Furthermore, one might argue that pre-Industrial Revolution civilizations are irrelevant in this context, as society has fundamentally changed—we are in a different reference class.<sup>82</sup>

## The Doomsday Argument

Another reason to think that near-term existential risk is high is the Doomsday Argument. First consider the *Self-Sampling Assumption*:<sup>83</sup>

**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:

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.

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

---

<sup>79</sup>Taagepera (1979, p. 133).

<sup>80</sup>Taagepera (1978, p. 191).

<sup>81</sup>Kemp (2019).

<sup>82</sup>Thank you Stefan Schubert for this point.

<sup>83</sup>The argument presented here is from Bostrom (2008). See also Carter (1974), Leslie (1996, §5) and Bostrom (2002) on the Doomsday Argument.

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 paper. 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**

The Vulnerable World Hypothesis is a thought experiment that concerns the potential risks associated with technological advancement.<sup>84</sup> The hypothesis proposes that there could be some level of technological development at which civilization almost certainly gets devastated by default, unless a sufficiently strong and widely implemented system of preventive policing or surveillance is in place.

**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.<sup>85</sup>

The idea is that as technology advances, the ability of individual actors to cause harm increases, making society more vulnerable. In Nick Bostrom’s words, some areas of science might make a discovery that ‘democratizes mass destruction.’<sup>86</sup>

The central idea is illustrated using a metaphorical ‘urn of invention.’<sup>87</sup> Each new idea, discovery and technology humanity produces is like drawing a ball from

---

<sup>84</sup>The Vulnerable World Hypothesis is from Bostrom (2019). See also von Neumann (1955).

<sup>85</sup>Bostrom (2019, p. 457).

<sup>86</sup>Bostrom (2019, p. 455).

<sup>87</sup>Bostrom (2019, p. 455).

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 mis-used that it could result in our destruction.

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.<sup>88</sup> It seems likely that social order would, at best, be severely disrupted by draconian measures to prevent disgruntled individuals from obtaining these materials. More pessimistically, societies might descend into civil wars fought with nuclear weapons, turning civilizations to ashes. Furthermore, if ‘civilization began to rise from the ashes, the knowledge would lie in wait, ready to pounce...’<sup>89</sup>

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.<sup>90</sup> The more people are 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.<sup>91</sup> I find the Vulnerable World Hypothesis the most plausible argument for relatively high near-term existential risk.

To summarize, I have discussed some reasons to think that existential risk is high: there is evidence for particular risks, species and civilizations have gone extinct, the Doomsday Argument and the Vulnerable World Hypothesis.

---

<sup>88</sup>Bostrom (2019, pp. 456–457).

<sup>89</sup>Bostrom (2019, p. 457).

<sup>90</sup>Kevin Kelly, in personal communication with Steven Pinker. See Pinker (2019, p. 302). See also Kelly (2012).

<sup>91</sup>Bostrom (2019, p. 458).

## 5 Final thoughts on the Optimistic Meta-Induction

Next, I will turn to some closing reflections on the Optimistic Meta-Induction.

### Objective chances

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.<sup>92</sup> Furthermore, while superforecasters tended to give much lower estimates for existential risks than subject matter experts, both groups nevertheless produced similar forecasts for non-anthropogenic risks.<sup>93</sup> It is reasonable to believe that these estimates are better grounded than those for anthropogenic risks, as it is generally easier to obtain evidence for the true rate of non-anthropogenic risks.

### 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. For example, Einstein, Oppenheimer, Russell and others may have been correct in pointing out that there was a great risk of a nuclear apocalypse. Luckily, though, we only experienced a few close calls. Nonetheless, we may set aside religious doomsday predictions, as they were not based on careful reasoning. 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

---

<sup>92</sup>Ćirković 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.

<sup>93</sup>Karger et al. (2023, p. 17).

our psychology—specifically, that we have a bias toward exaggerating existential risks.

### **Existential risks possible**

When 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**

As I noted at the beginning of this paper, 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.<sup>94</sup> So, even small decreases in overall existential risk could lead to enormous gains in expected moral value.<sup>95</sup> 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.”

## **6 Conclusion**

Many assign relatively high probabilities to near-term existential risks. However, our estimates of existential risk probabilities might be undermined by the Opti-

---

<sup>94</sup>Bostrom (2013).

<sup>95</sup>Bostrom (2013). For a contrary view, see Thorstad (2023).

mistic 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 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, I conclude that we should be careful before assigning a high subjective credence to near-term existential risks.<sup>96</sup>

## References

Avenhaus, R., Fichtner, J., Brams, S. J. and Kilgour, D. M. (1989), ‘The probability of nuclear war’, *Journal of Peace Research* **26**(1), 91–99.

Aveni, A. (2016), *Apocalyptic Anxiety: Religion, Science, and America’s Obsession with the End of the World*, University Press of Colorado, Boulder, Colorado.

Bailey, R. (2000), ‘Earth day, then and now’.

URL: <https://reason.com/2000/05/01/earth-day-then-and-now-2/>

Barrett, A. M. (2016), ‘False alarms, true dangers? current and future risks of inadvertent U.S.-Russian nuclear war’, <https://nsarchive2.gwu.edu/NSAEBB/NSAEBB75/>. RAND Corporation document PE-191-TSE.

Barrett, A. M., Baum, S. D. and Hostetler, K. (2013), ‘Analyzing and reducing the risks of inadvertent nuclear war between the united states and russia’, *Science & Global Security* **21**(2), 106–133.

Baum, S. (2015), ‘Breaking down the risk of nuclear deterrence failure’, <https://thebulletin.org/2015/07/>

---

<sup>96</sup>Johan Gustafsson suggested that we should not believe the conclusion of this paper because most past philosophical papers have turned out to be wrong.

breaking-down-the-risk-of-nuclear-deterrence-failure/. Bulletin of the Atomic Scientists.

Baum, S., de Neufville, R. and Barrett, A. (2018), 'A model for the probability of nuclear war', Global Catastrophic Risk Institute Working Paper 18-1.

**URL:** <https://ssrn.com/abstract=3137081>

Bereanu, B. (1983), 'Self-activation of the world nuclear weapons system', *Journal of Peace Research* **20**(1), 49–57.

Bird, K. and Sherwin, M. J. (2005), *Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer*, Alfred A. Knopf, New York.

Blanton, T. (1997), 'Annals of blinksmanship', [https://nsarchive2.gwu.edu/nsa/cuba\\_mis\\_cri/annals.htm](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.

Borrie, J. (2014), 'A limit to safety: Risk, 'normal accidents', and nuclear weapons'. ILPI-UNIDIR Vienna Conference Series, paper 3.

Bostrom, N. (2002), *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, Routledge, 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.

- Boyle, F. A. (1985), *World Politics and International Law*, Duke University Press, Durham, NC.
- Brito, D. L. and Intriligator, M. D. (1996), ‘Proliferation and the probability of war: A cardinality theorem’, *Journal of Conflict Resolution* **40**(1), 206–214.
- Burr, W. and Blanton, T. S. (2002), ‘The submarines of October: U.S. and Soviet naval encounters during the Cuban missile crisis’, <https://nsarchive2.gwu.edu/NSAEBB/NSAEBB75/>. National Security Archive Electronic Briefing Book No. 75.
- Carter, B. (1974), ‘Large number coincidences and the anthropic principle in cosmology’, pp. 291–298.
- Chakravartty, A. (2017), Scientific Realism, in E. N. Zalta, ed., ‘The Stanford Encyclopedia of Philosophy’, Summer 2017 edn, Metaphysics Research Lab, Stanford University.
- Ćirković, M. M. (2008), Observation selection effects and global catastrophic risks, in N. Bostrom and M. M. Ćirković, eds, ‘Global Catastrophic Risks’, Oxford University Press, Oxford, pp. 120–145.
- Ćirković, M. M., Sandberg, A. and Bostrom, N. (2010), ‘Anthropic shadow: Observation selection effects and human extinction risks’, *Risk Analysis* **30**(10), 1495–1506.
- Conn, A. (2016), ‘Accidental nuclear war: a timeline of close calls’, Future of Life Institute.  
**URL:** <https://futureoflife.org/resource/nuclear-close-calls-a-timeline/>
- Dixon, B. (1971), ‘In praise of prophets’, *New Scientist* p. 606.  
**URL:** <https://archive.org/details/screenshot-of-ehrlachs-claim-that-england-will-not-exist>

- Ebell, M. and Milloy, S. J. (2019), 'Wrong again: 50 years of failed eco-pocalyptic predictions.'  
**URL:** <https://cei.org/blog/wrong-again-50-years-of-failed-eco-pocalyptic-predictions/>
- 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.
- 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/>
- Gross, M. B. and Gilles, M. (2012), *The Last Myth: What the Rise of Apocalyptic Thinking Tells Us about America*, Prometheus Books, Amherst, NY.
- Hall, J. R. (2009), *Apocalypse: From Antiquity to the Empire of Modernity*, Polity Press, Cambridge, UK.
- Hellman, M. E. (2008), 'Risk analysis of nuclear deterrence', *The Bent of Tau Beta Pi* pp. 14–22.
- Hoffman, D. E. (2011), *The Dead Hand: Reagan, Gorbachev and the Untold Story of the Cold War Arms Race*, Icon Books, London.
- Huchthausen, P. A. (2002), *October Fury*, John Wiley & Sons, Hoboken, New Jersey.

- Hughes, J. J. (2008), Millennial tendencies in responses to apocalyptic threats, *in* N. Bostrom and M. M. Ćirković, eds, 'Global Catastrophic Risks', Oxford University Press, Oxford, pp. 73–90.
- Intriligator, M. D. and Brito, D. L. (1981), 'Nuclear proliferation and the probability of nuclear war', *Public Choice* **37**(2), 247–260.
- 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>
- Laudan, L. (1981), 'A confutation of convergent realism', *Philosophy of Science* **48**(1), 19–49.
- Leslie, J. (1996), *The End of the World: The Science and Ethics of Human Extinction*, Routledge, London.
- Lewis, P., Williams, H., Pelopidas, B. and Aghlani, S. (2014), 'Too close for comfort: Cases of near nuclear use and options for policy'. Chatham House Report.
- Lundgren, C. (2013), 'What are the odds?: Assessing the probability of a nuclear war', *The Nonproliferation Review* **20**(2), 361–374.
- Malthus, T. (1798), *An Essay on the Principle of Population: An Essay on the Principle of Population, as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers*, J. Johnson, London.

- McGovern, J. R. (1992), *The World of Columbus*, Mercer University Press, Macon, Georgia.
- McIver, T. (1999), *The End of the World: An Annotated Bibliography*, McFarland & Company, Jefferson, North Carolina.
- 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.
- Mozgovoi, A. (2002), 'The Cuban samba of the quartet of foxtrots: Soviet submarines in the Caribbean crisis of 1962', <http://nsarchive.gwu.edu/NSAEBB/NSAEBB75/asw-II-16.pdf>. Military Parade, Moscow. Translated by Svetlana Savranskaya, The National Security Archive.
- 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.
- Psillos, S. (2022), Realism and theory change in science, in E. N. Zalta and U. Nodelman, eds, 'The Stanford Encyclopedia of Philosophy', Fall 2022 edn, Metaphysics Research Lab, Stanford University.

- Randle, M. J. and Eckersley, R. (2015), 'Public perceptions of future threats to humanity and different societal responses: A cross-national study', *Futures* 72, 4–16.
- 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.
- Revelation (2017), *The Holy Bible*, Holman Bible Publishers, Nashville, Tennessee.
- Russell, B. (1951), 'The future of man', *Atlantic* .
- Sagan, S. D. (1993), *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*, Princeton University Press, Princeton, New Jersey.
- Sagan, S. D. and Waltz, K. N. (1995), *The Spread of Nuclear Weapons: A Debate*, W. W. Norton & Company, New York.
- Savranskaya, S. V. (2005), 'New sources on the role of Soviet submarines in the Cuban missile crisis', *Journal of Strategic Studies* 28(2), 233–259.
- 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.
- Taagepera, R. (1978), 'Size and duration of empires growth-decline curves, 3000 to 600 b.c.', *Social Science Research* 7(2), 180–196.
- Taagepera, R. (1979), 'Size and duration of empires: Growth-decline curves, 600 b.c. to 600 a.d.', *Social Science History* 3(3/4), 115–138.

- Tegmark, M. and Bostrom, N. (2005), 'Is a doomsday catastrophe likely?', *Nature* **438**(7069), 574.
- 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.
- Union of Concerned Scientists (2015), 'Close calls with nuclear weapons', <https://www.ucsusa.org/sites/default/files/attach/2015/04/Close%20Calls%20with%20Nuclear%20Weapons.pdf>.
- von Neumann, J. (1955), 'Can we survive technology?', *Fortune* pp. 106–152.
- Walker, J. (2020), 'Trump told Woodward 'you don't know how close we were to war' with North Korea'.  
**URL:** <https://www.newsweek.com/trump-woodward-dont-know-how-close-war-north-korea-1531634>
- 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.
- 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](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](https://ygo-assets-websites-editorial-emea.yougov.net/documents/tabs_OPI_disaster_20150227_QaIqpU3.pdf).
- YouGov (2023), 'Yougov survey: AI and the end of humanity', [https://d3nk13psvxxpe9.cloudfront.net/documents/results\\_AI\\_and\\_the\\_End\\_of\\_Humanity.pdf](https://d3nk13psvxxpe9.cloudfront.net/documents/results_AI_and_the_End_of_Humanity.pdf).
- Yudkowsky, E. (2008), Cognitive biases potentially affecting judgement of global risks, *in* N. Bostrom and M. M. Čirković, eds, 'Global Catastrophic Risks', Oxford University Press, Oxford, pp. 91–119.