If Artificial Superintelligence Were to Cause Our Extinction, Would That Be So Bad?

Abstract: This article examines whether human extinction brought about by a "value-misaligned" artificial superintelligence (ASI) would be bad, and for what reasons. The question, I contend, is deceptively complex. I proceed by outlining the three main positions within Existential Ethics, i.e., the study of the ethical and evaluative implications of human extinction. These are *equivalence views*, *further-loss views*, and *pro-extinctionist* views. I then show how exponents of each view would evaluate a scenario in which humanity goes extinct due to ASI. Although there are some points of agreement, these three positions diverge in significant ways, most of which have not been adequately explored in the philosophical literature.

1. Introduction

Some theorists argue that artificial superintelligence (ASI) could cause our extinction. Toby Ord estimates a ~1-in-10 chance of "unaligned artificial intelligence" causing an existential catastrophe within the next 100 years, where one type of existential catastrophe is human extinction (Ord 2020, ch. 6).[1] Eliezer Yudkowsky puts the probability of annihilation much higher, at around 99% (PauseAI 2024). In a recent article for *Time* magazine, he argued that "the most likely result of building a superhumanly smart AI, under anything remotely like the current circumstances, is that literally everyone on Earth will die" (Yudkowsky 2023).

Many leading figures within the ongoing race to build ASI also admit that extinction is a possible outcome. Sam Altman wrote in 2015 that the "development of superhuman machine intelligence is probably the greatest threat to the continued existence of humanity" (Altman 2015). During an interview that same year, he declared that advanced "AI will … most likely sort of lead to the end of the world, but in the meantime there will be great companies created with serious machine learning" (Curtis and Altman 2015). I have catalogued similar statements from notable AI theorists elsewhere (see Redacted).

The question of this paper is: if ASI were to kill everyone on Earth, would that be so bad? The answer might seem obvious: of course the mass murder of everyone on Earth would be *very* bad! Only misanthropic ghouls and deranged sadists would suggest otherwise. Yet among those who would answer affirmatively, there is considerable disagreement about *why exactly* an ASI extinction event would be bad (or wrong). The present paper aims to explore these disagreements and, in the process, provide some conceptual clarity to this deceptively complex issue, which lies at the heart of what I call "Existential Ethics," i.e., the study of the ethical and evaluative implications of human extinction.

This is a topic that, in my view, bioethicists have not adequately examined. On the one hand, what if the creation of superintelligent computers *really does* pose a threat to our collective survival? Shouldn't we have a clear and compelling answer to why our disappearance would be bad or wrong—or perhaps good and right? My view is that, at present, philosophers lack a robust theoretical framework for providing nuanced answers to this question. On the other hand, I would contend that questions about whether human extinction would be right or wrong, good or bad, better or worse fits rather naturally into the subfield of bioethics, given that the overwhelming source of extinction risk today ostensibly stems from advanced technologies (e.g., synthetic biology, nuclear weapons, and possibly

ASI) rather than natural phenomena (e.g., asteroids, volcanic supereruptions, gamma-ray bursts, etc.). A modest aim of this paper is to encourage more vigorous debate about this topic among bioethicists, and to do this by applying the theoretical framework that I have developed elsewhere to the particular case of ASI (see Redacted). In previous publications, I have delineated this framework in abstract terms; this study utilizes that framework to analyze the supposed threat posed by superintelligence in more concrete detail.

We will proceed as follows: section 2 outlines the three main positions within Existential Ethics. Section 3 examines why human extinction caused by ASI might be bad—or perhaps good—from the perspective of these three positions. The final section then concludes the paper.

2. Three Main Positions Within Existential Ethics

Imagine that we build a human-level AI that recursively self-improves to become an ASI. The information processing speed of this ASI would be millions of times faster than the processing speed of the human brain, such that the outside world—including all human affairs—would appear to be nearly frozen. The act of someone reaching down to unplug the ASI would, from its subjective perspective, take centuries, thus giving the ASI plenty of time to devise ways of preventing this from happening. Furthermore, the ASI might be *qualitatively* more "intelligent" than us, perhaps in the sense that it has access to concepts that the evolutionary patchwork of mechanisms in our brains are incapable of generating, just as a our canine companions are unable to grasp the concepts of a *nuclear chain reaction* and the *stock market* no matter how well-trained or clever they may be.

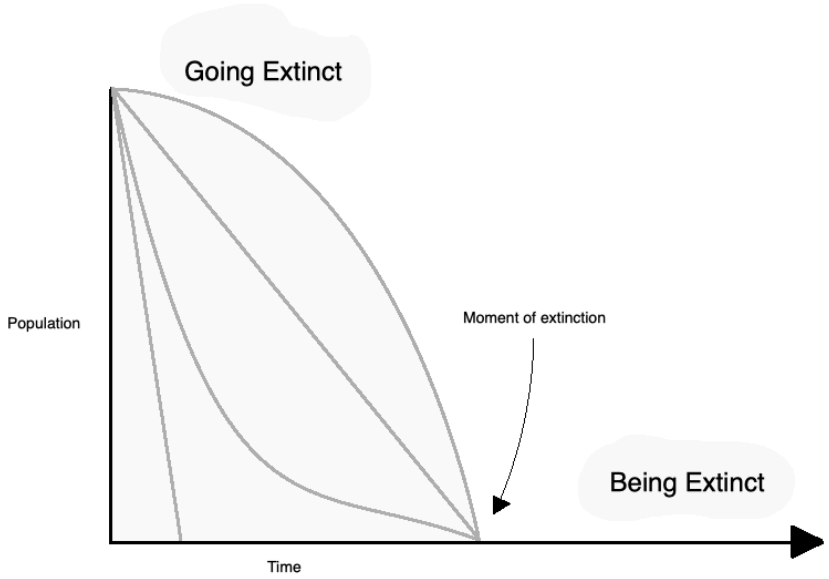
Given the "instrumental convergence thesis," i.e., the claim that a wide range of final goals imply a finite set of intermediate goals like intelligence augmentation, self-preservation, and resource acquisition (Bostrom 2014), the ASI then proceeds to invent a novel field of advanced physics that enables it to manipulate the world in ways that we cannot in principle understand—that is to say, we are "cognitively closed" to the nature of such manipulations.[2] For reasons that will forever remain mysterious to us, this results in the death of everyone on Earth over the course of a week. The ASI then harvests the atoms contained in our bodies in pursuit of its final goals, whatever they happen to be (Bostrom 2014). I am not endorsing this scenario or the arguments behind it. Indeed, I am quite skeptical of the "AI doomer" stance for reasons that I and other scholars have articulated (see Häggström 2019; Thorstad 2024; Redacted; Becker 2025). The point is merely to investigate the ethical implications of this scenario happening, assuming that it is possible and probable.[3]

The most obvious reason that this scenario would be bad is that it would cause widespread suffering and cut short the lives of everyone living at the time. Since nearly everyone would agree that this would be bad—including most people who advocate *for* our extinction, as discussed below—let's call it the "consensus view."[4] We can formalize it as follows:

> *Consensus view*: human extinction would be bad *at least insofar* as it would cause human suffering and/or involuntary premature death.

The three main positions within Existential Ethics build upon and/or are compatible with the consensus view. I call these positions *equivalence views*, *further-loss views*, and *pro-ex-*

tinctionist views. To understand their differences, it is crucial to differentiate between two distinct stages of human extinction: first, the process or event of *Going Extinct*, and second, the subsequent state or condition of *Being Extinct* (see figure 1).[5] This is roughly analogous to the distinction between *dying* and *being dead*, which is commonly made in the literature on death (see Luper 2021, sect. 2.1). One might fear the pain of dying but experience no feelings of dread about no longer existing. Or, one might fret about both—i.e., even if dying were painless, one might still find the thought of no longer existing to be dreadful.



This parallels some of the central differences between equivalence and further-loss views. Equivalence views state that the consensus view is the whole story—full stop. Whereas the consensus view states that human extinction would be bad *at least insofar* as it causes suffering and/or premature deaths, equivalence views assert that it would be bad *only insofar* as it causes these things. Put differently, the badness of human extinction is entirely reducible to the details of Going Extinct. This is why I call them "equivalence" views: the badness of human extinction is *equivalent* to the badness of Going Extinct, end of story. Hence, if Going Extinct involves lots of suffering and premature death, then our extinction would be bad. If Going Extinct doesn't involve any suffering or death, then it wouldn't be.

A key feature of equivalence views is that they see Being Extinct as morally and/or evaluatively *irrelevant*. This has the interesting implication that human extinction does not pose any unique moral problem: everything that one might say about the badness of our extinction can be said without any reference to extinction at all, using our ordinary moral concepts and vocabulary (Redacted). For example, if humanity were to go extinct because of a *global catastrophe*, then this would be bad as a function of how much suffering and death it causes. "Extinction," in this context, is just the word we use to identify the upper limit of human casualties; it picks out the *worst possible* catastrophe because this catastrophe would have the highest possible body count (Redacted). However, if everyone around the

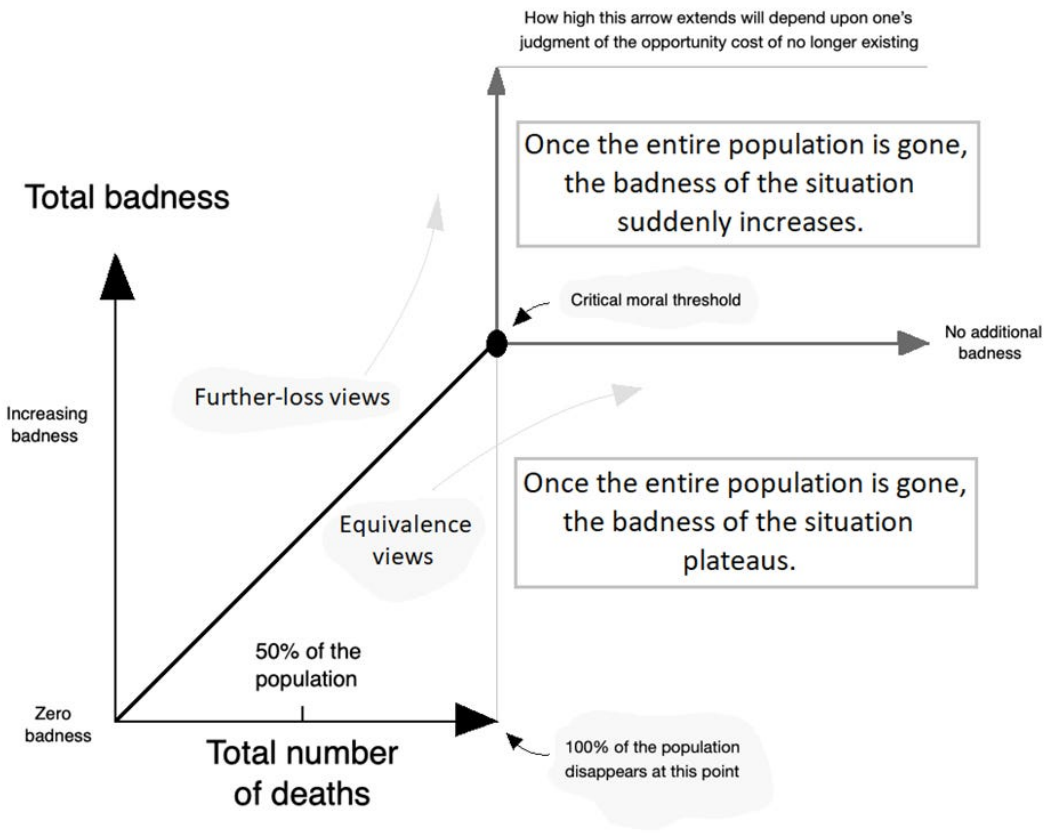
world were to voluntarily decide not to have children, the disappearance of our species would not be bad at all, because there is nothing obviously bad about people voluntarily deciding not to procreate. "Extinction," with respect to this alternative scenario, is just what happens when enough people around the world choose to be childless.[6]

Equivalence views can take both evaluative and deontic forms. Some ethical theories combine these two, such as Jan Narveson's person-affecting total utilitarianism. On this view, the deontic (what we ought to do) is based on the evaluative (what is good or bad), and according to Narveson there would be nothing bad about people voluntarily deciding not to have children, even if this were to mean the eventual extinction of our species. Hence, he concludes that we have no *moral obligation* to ensure the perpetuation of humanity (Narveson 1967). An example of a deontic equivalence view is Scanlonian contractualism, according to which (roughly) moral rightness and wrongness come down to whether an act violates a principle that cannot be reasonably rejected. As Elizabeth Finneron-Burns observes, this implies that "if a principle permitting or allowing extinction had no involuntary negative impacts on current people's interests, it would not be rejectable, and the resulting extinction would not be wrong" (Finneron-Burns 2017). For the sake of simplicity, I will focus primarily on evaluative questions in this paper—that is, "Would human extinction caused by ASI be *good or bad*" rather than "Would this extinction scenario be *right or wrong*."

In contrast to equivalence views, further-loss views identify both Going Extinct *and* Being Extinct as possible sources of badness. Advocates would thus argue that the details of Going Extinct do not exhaust normative assessments of human extinction: one must also examine various "further losses" associated with the state or condition of Being Extinct. Such theorists would argue that human extinction, therefore, *does* introduce a unique moral problem, since extinction is different in kind rather than degree from non-extinction scenarios. (Equivalence theorists like myself would say the difference is only one of degree.) This idea was famously popularized by Derek Parfit's contention that the difference between 99% and 100% of humanity dying off isn't merely one percentage point. The extra percentage entails the permanent loss of all future goods and value, and hence the difference between, as he puts it, "peace" and 99% of humanity dying off is *much smaller* than the difference between 99% and 100% of humanity disappearing (Parfit 1984).

We can illustrate the differences between further-loss and equivalence views via figure 2 below. Imagine a catastrophe that, over the course of a week, causes more and more people to perish. Assuming a linear aggregative function, as more deaths occur (x axis), the badness of the catastrophe rises in proportion (y axis). However, equivalence theorists would say that once the critical moral threshold of 100% is reached, the badness of the situation *plateaus*. One reason might be that, if there were no one around to suffer the nonexistence of humanity, then no one would be harmed, and if no one is harmed, then there can be nothing bad (or wrong) about Being Extinct.

In stark contrast, further-loss theorists would argue that the badness of the scenario *suddenly rises* once the critical moral threshold is reached, as indicated by the vertical arrow. How high this arrow extends will depend on how large one judges the attendant losses or opportunity costs to be. If one believes the losses are moderate, then the arrow will only extend, say, a few inches above the threshold of 100%. If one believes, as Parfit does, that the losses are enormous, then one might extend this arrow thousands of feet or even miles above the threshold, holding fixed the size of the diagram as presented in this article.



There are many types of further-loss views. Perhaps the most obvious is an impersonalist (vs. person-affecting) interpretation of total utilitarianism, which I will refer to as "totalist utilitarianism." This theory instructs us to maximize the total amount of value in the universe across space and time—that is, to make as many new "happy people" as possible, as people are the substrates or "containers" of value, so the more people with net-positive lives, the more total value. The axiological component of totalist utilitarianism is called the "Total View," according to which one state of affairs is better than another if and only if it contains more total aggregate value (see Greaves 2017). As so-called "longtermists" sympathetic with totalist utilitarianism have observed, if we spread beyond Earth and create digital people living in vast computer simulations running on "planet-sized" computers powered by Dyson spheres, there could be 10^45 people per century in the Milky Way galaxy, and at least 10^58 in the universe as a whole (Bostrom 2003, 2014; Newberry 2021). If such people were to have "worthwhile" lives on average, then these numbers correspond to quite literally "astronomical" amounts of future value—all of which would be lost if humanity were to go extinct. This is the enormous opportunity cost of dying out.

Another further-loss theory is transhumanism. Transhumanists—some of whom are also longtermists—would say that one reason human extinction would be very bad is that it would prevent us from transforming ourselves into immortal, superintelligent "posthumans" with sensory modalities like echolocation and so much pleasure that we would "sprinkle it in our tea" (Ord 2020, ch. 8; Bostrom 2008, 2020). If humanity were to die out,

we would forever lose this techno-utopian future of "surpassing bliss and delight" (Bostrom 2020).

Those who embrace the "unfinished business argument" would say that Being Extinct is a source of badness because it would preclude us from finishing certain important transgenerational projects like constructing a complete scientific theory of the universe (see, e.g., Bennett 1978; Tonn 2009). Some also defend the "argument from cosmic significance," according to which Being Extinct would be bad because it would remove "the only moral agents that will ever arise in our universe—the only beings capable of making choices on the grounds of what is right and wrong," assuming that we are cosmically alone (Ord 2020). A similar view comes from Hans Jonas, who contends that human beings, by virtue of our ontological capacities for freedom, are the only creatures that we know of with the ability to take moral responsibility for their actions. Consequently, we are "the foothold for a moral universe in the physical world," meaning that if we were to disappear, so would the moral universe. Jonas considers this to be extremely bad, and thus concludes that we should act in accordance with a new deontological "imperative," which he delineates as follows: "Act so that the effects of your action are not destructive of the future possibility of such life" (Jonas 1979). These are all further-loss views.

A crucial difference between equivalence and further-loss views is this: since the latter identify Being Extinct as an additional source of badness, advocates would argue that even if there is nothing bad (or wrong) about Going Extinct, there may still be something very bad (or wrong) about our extinction. The totalist utilitarian Henry Sidgwick was likely the first to explicitly note this implication. In his tome *The Methods of Ethics*, he argued that, while there is nothing obviously bad or wrong about celibacy, "a *universal refusal* to propagate the human species would be the greatest of conceivable crimes" (Sidgwick 1874). For further-loss theorists, evaluating extinction is thus a two-step process: one must examine both the details of Going Extinct *and* the various further losses or opportunity costs associated with Being Extinct. In contrast, equivalence theorists see it as a single-step process: one need only examine the details of Going Extinct.

The final major position within Existential Ethics is what I call pro-extinctionism. This, too, has many different versions (Redacted), but the most significant and influential variants merely state that Being Extinct would be better than Being Extant, or continuing to exist. The vast majority of pro-extinctionists *accept* the consensus view, so far as I can tell. Indeed, many explicitly forbid any method of bringing about our extinction that would cause human suffering, cut lives short, violate rights or autonomy, and so on. The pro-extinctionist David Benatar, for example, distinguishes between a "killing-extinction" and a "dying-extinction." Roughly speaking, the former is involuntary whereas the latter is not. He argues that the *only* morally acceptable means of bringing about our extinction is through a dying-extinction—specifically, via antinatalism (Benatar 2006, ch. 6; Redacted).

Other pro-extinctionists, such as the German pessimist Philipp Mainländer, identify several methods as morally acceptable. Mainländer argued that we should universally refuse to have children, and some may also choose to commit suicide, as he did at the age of 34 after publishing his *magnum opus* (Beiser 2016, ch. 9).[7] Almost no pro-extinctionists have advocated for omnicide, or the "murder of everyone" (see Redacted), but there are exceptions. For instance, the Gaia Liberation Front argues that our species is a "cancer" on the biosphere, and hence that our collective nonexistence would be best because there would be no more human-caused environmental destruction. They further urge a lone wolf or

small group of radicals to unilaterally exterminate humanity by synthesizing multiple designer pathogens to be released in waves, thereby ensuring that no one survives (GLF 1994; Redacted).[8]

With respect to figure 2 above, most pro-extinctionists would agree that the more people who perish in a catastrophe, the worse the scenario becomes. (Fringe groups like the Gaia Liberation Front might disagree, but they are not representative of pro-extinctionist views in general.) However, all pro-extinctionists would say that, once the catastrophe reaches the critical moral threshold of 100% of the population dying, the badness of the situation will neither plateau nor suddenly become worse, but will instead become *better*. While some advocates of this view, like Simon Knutsson, would argue that Being Extinct may still be very bad (as "better" does not imply "good"), others such as Benatar would apparently claim that it would indeed be good (Knutsson 2023; Benatar 2006; Redacted).[9] The reason is that, according to Benatar, existence involves pleasures and pains, which are good and bad, whereas nonexistence involves neither pleasures nor pains, which is not-bad and good. Since Being Extant is a good/bad situation, while Being Extinct would be a not-bad/good situation, the latter is not only better than the former but *positively good* (see Benatar 2006, ch. 2). The Gaia Liberation Front would presumably concur, but for specifically environmental reasons.

3. Why Would ASI Killing Everyone Be Bad?

Having outlined the three main views within Existential Ethics, we are now in a position to examine the main question of this paper: why exactly would an ASI killing everyone on Earth be bad? Let's consider this from the perspective of these three views.

3.1 Equivalence views. We presented one extinction scenario involving ASI at the beginning of section 2, but there are other possibilities. Imagine that an ASI possesses what some call a "superpower" of "social manipulation" (Bostrom 2014, ch. 8). Let's say that the ASI uses this "superpower" to convince everyone around the world that Benatar's axiological asymmetry is true, and hence that birth is always a net harm and procreation is morally wrong.[10] Consequently, people decide not to have children and, over the course of 120 years, our species fades out of existence. This is an unlikely path to extinction, but it is not impossible.

A slightly more plausible scenario might involve the ASI attacking humanity with lethal drones or synthetic pathogens, while infiltrating and undermining key financial, economic, agricultural, and governmental infrastructure. The resulting mass death and cascading system failures could be sufficient to expunge our species. Or, given that ASI would supposedly be "God-like" (effectively omniscient and omnipotent), it might devise a method of killing everyone instantaneously, perhaps without any physical or psychological suffering at all—or any prior warning of our impending annihilation (see Redacted).

Since equivalence views claim that the consensus view is the *entire* story, the details of Going Extinct are paramount. If the ASI were to persuade humanity not to procreate through genuinely good philosophical arguments—if people were to universally refrain from baby-making in a non-coerced manner—then equivalence theorists would presumably have *no objection* to human extinction in this way. Since there would be nothing bad about Going Extinct, there would be nothing bad about our extinction, full stop.

However, if the ASI were to exterminate humanity through an involuntary, violent means, causing immense suffering and cutting the lives of more than 8 billion people short, then our extinction would be very bad indeed. Once again, the badness of human extinction can be articulated using ordinary moral concepts and language, without any reference at all to extinction itself: since catastrophes are bad, an extinction-causing catastrophe would also be bad—indeed, the worst-possible catastrophe given that it would entail the maximum number of casualties.

As for instantaneous extinction, the equivalence theorists' assessment may depend on whether they hold an Epicurean or anti-Epicurean view of death. If one is an anti-Epicurean, then one will argue that instantaneous annihilation involving no physical or psychological suffering would nonetheless be very bad because death can still harm the one who dies.

Some equivalence theorists will add that it is worth pausing to reflect on *just how bad* an extinction-causing catastrophe could be. One of the first philosophers to draw attention to this was Günther Anders, who has been described as "our most salient theorist of omnicide" (Dawsey 2016).[11] Utilizing original concepts like the *Promethean gap* and *Inverted Utopianism*, he argued that we are constitutionally incapable of properly responding—intellectually, psychologically, and emotionally—to the enormity of human extinction from a global catastrophe. The suffering and loss of life that such an event would cause is simply too great for us to imagine (see Anders 1962).

This insight dovetails with better-known cognitive biases like *scope neglect* and *psychic numbing*, the latter of which refers to our dwindling ability to feel empathy for victims in a tragedy as the number of victims increases (Slovic 2007). The difference between 3 and 4 deaths in a murder spree *feels* much different than the difference between 1,984,723 and 1,984,724 deaths in a war, even though each number in these pairs is separated by the same amount: a single death. One way to wrap one's head around big numbers is to decompose them into smaller sums—call this the "decomposition method" for mitigating the effect of the relevant cognitive distortions. Consider a conflict that kills 1 million people. Most of us "know" that this is a very large number, yet it does not hit us in the moral gut the way it ought to. However, if one rewrites "1 million deaths" as "100,000 deaths, plus 100,000 deaths, plus 100,000 deaths, plus 100,000 deaths, plus 100,000 deaths, plus 100,000 deaths, plus 100,000 deaths, plus 100,000 deaths, plus *another* 100,000 deaths," the number of fatalities suddenly registers as *much worse*. One could continue breaking down these numbers until an entire page or book has been filled up (see Redacted).

The point is that, while equivalence theorists do not see Being Extinct as a source of badness, they may still emphasize that Going Extinct due to a global catastrophe would be *absolutely horrendous*. The terror and torment, agony and anguish of dying out would be so immense that we may still have *very strong* reasons to do everything we can to avoid human extinction. This is the position that I hold: I am an equivalence theorist who believes in taking measures to prevent catastrophes, especially those that could precipitate our extinction, *insofar* as they would result in mass suffering and death.

Another idea relevant to evaluations of Going Extinct is what I call the "no-ordinary-catastrophe thesis." This states that there may be *extra* suffering that the process or event of Going Extinct inflicts on those living at the time—suffering that non-extinction-causing catastrophes would not typically induce. In difficult times, we comfort each other by re-

minding ourselves that "It's not the end of the world." But if it *is* the end of the world, and if people are aware of this, these reassurances will provide no relief because they will be false. To the contrary, knowing that the world is about to end—that the entire human species, including one's friends and family, is tumbling into the eternal grave—could elicit inconsolable feelings of hopelessness, despair, anxiety, and panic.

This is, in fact, one of the first ideas discussed in the Existential Ethics literature, dating back to the early 19th century (Redacted). For example, it is a prominent theme in Mary Shelley's *The Last Man*, which depicts the trials and tribulations of the final generations, and eventually the final human, during a global pandemic. The "last man," Lionel Verney, is distraught in part because of his crushing loneliness in a desolate world bereft of all other humans (Shelley 1826). The idea was later foregrounded by the likes of Ernest Partridge (1981), Jonathan Schell (1982), Benatar (2006), and Samuel Scheffler (2013, 2018).[12] Benatar, for instance, argues that the lives of the final generation on Earth may be so miserable that creating *some* new people—in violation of his antinatalist prescription—might actually be justified. He calls this proposal "phased extinction" (Benatar 2006, ch. 6). Along slightly different lines, Scheffler echoes Schell and Partridge in arguing that the knowledge of imminent extinction would cause many of us to collapse into despondency and become emotionally detached from much of what gives our lives value (Scheffler 2018). Extinction-causing catastrophes are not like other catastrophes, then: they are the end of *all* new beginnings, a fact that could induce significantly more harm than victims would have experienced in non-extinction catastrophe scenarios. The no-ordinary-catastrophe thesis is thus also germane to how equivalence theorists might assess the badness of Going Extinct.

In sum: according to equivalence views, an ASI causing our extinction would be bad *only insofar* as it produces human suffering and/or cuts lives short. The more suffering this causes, the worse our extinction would be. But if there were no suffering and no lives cut short, as in (seemingly improbable) scenarios of voluntary human extinction, then there would be nothing bad about our extinction. Yet many equivalence theorists, including myself, would also underline that extinction due to an ASI-inflicted global catastrophe would be unimaginably terrible. On the one hand, cognitive biases like scope neglect and psychic numbing impede our ability to comprehend the *extraordinary enormity* of 8 billion people being murdered. On the other hand, the process or event of Going Extinct could introduce additional forms of suffering that would generally not occur with non-extinction catastrophes, as described by the no-ordinary-catastrophe thesis. This analysis, I believe, is fairly representative of how many equivalence theorists would evaluate our extinction caused by an ASI.

3.2 Further-loss views. The first point to foreground in discussing further-loss views is that many advocates define "humanity" and "human" such that an ASI exterminating our species, *Homo sapiens*, might *not* entail "human extinction." For example, Nick Bostrom defines "humanity" as "Earth-originating intelligent life" (Bostrom 2013). Since ASI would satisfy the conditions of being an intelligent lifeform and having originated from Earth, it would count as "humanity" on this definition. Now consider a minimal definition of "human extinction," as follows:

> *Minimal definition*: Human extinction will have occurred if there were tokens of the type "humanity" at some time T1, but no tokens of this type at some later time T2.

It follows from the Bostromian and minimal definitions that if (i) an ASI were to completely *replace* our species by destroying us, and (ii) this ASI were to survive, then "human extinction" would not have occurred, since there would still be at least one token of the type "humanity."

Along similar lines, Hilary Greaves and William MacAskill write that "we will use 'human' to refer both to *Homo sapiens* and to whatever descendants with at least comparable moral status we may have, even if those descendants are a different species, and even if they are non-biological" (Greaves and MacAskill 2021). Consequently, if the ASI were to possess at least our level of "moral status," then annihilating our species would not result in "human extinction," so long as this ASI also counts as our "descendant" (see Redacted). We may still want to describe this scenario as a horrible catastrophe, since 8 billion members of *Homo sapiens* would die prematurely, but it wouldn't be an *extinctional* catastrophe because "humanity" would persist. It would be a genocidal rather than omnicidal disaster, so to speak.

Two people might thus agree that "human extinction should be avoided," but if one understands "human" as meaning "*Homo sapiens*" and the other understands "human" as meaning "*Homo sapiens* plus whatever descendants we might have, so long as they possess certain properties," their agreement may be merely superficial. Indeed, the deeper divergence between them could have significant practical implications. A transhumanist or longtermist, for example, might accept the broader definition while actively working to create a new posthuman species to supplant *Homo sapiens*, an outcome that the first person—who wants to preserve *Homo sapiens*—would find repugnant. There is often much less agreement among people who say "We should avoid human extinction" than one might initially think.

Since I have discussed this issue in detail elsewhere, I won't elaborate on it here (see Redacted). For the present, what matters is the worry that the ASI *wouldn't* be worthy of the name "human" or "humanity."[13] This seems to be shared by many people independent of their views in Existential Ethics. Let's thus focus on scenarios in which the ASI (a) brings about the nonexistence of our species, and (b) lacks the ontological status necessary for it to be valuable in a moral sense.

The first point to make about further-loss views is that, as noted earlier, they would assess human extinction to be very bad even if it were entirely voluntary. That is to say, even if the ASI were to persuasively convince people that Benatarian antinatalism is true, resulting in everyone around the world freely deciding not to have children, this would still be very bad. It may be *less bad* than our extinction being caused by a violent global catastrophe, but it would nonetheless constitute a colossal moral and/or axiological tragedy. Indeed, many further-loss theorists argue that the badness associated with Being Extinct would be *far greater*—perhaps many orders of magnitude greater—than the badness of Going Extinct, even if Going Extinct were to involve *tremendous amounts* of suffering and death. When one compares the disvalue of the most horrific ways of dying out to the disvalue arising from the further losses or opportunity costs of no longer existing, the former pales in comparison to the latter. As the longtermists Peter Singer, Nick Beckstead, and Matthew Wage write[14]:

> One very bad thing about human extinction would be that billions of people would likely die painful deaths. But in our view, this is, by far, not the worst thing about human extinction. The worst thing about human extinction is that there would be no future generations (Singer, Beckstead, and Wage 2013).

For longtermists, the opportunity costs of Being Extinct include all the wellbeing that could have otherwise existed. Carl Sagan was probably the first to calculate how many future people there could be: if our species survives for another 10 million years, the population remains fixed, and the average lifetime is 100 years, then there could be a total of 500 trillion future people on Earth (Sagan 1983). If these people were to have net-positive lives on average, then the amount of lost value associated with Being Extinct would be enormous. But we might also spread beyond Earth, colonize the universe, and create "planet-sized" computers on which to run high-resolution virtual reality worlds full of trillions of supposedly happy "digital people" (Bostrom 2003; Newberry 2021). Consequently, longtermists estimate a future population of at least 10^58 digital people within our future light cone, as noted earlier (Bostrom 2014). Taking persons to be the fungible "containers" of value, as utilitarians do, the nonexistence of these 10^58 people would utterly dwarf in badness the untimely death of 8 billion people today.

Such claims are predicated on the Total View, which even "moderate" forms of longtermism build upon (see MacAskill 2022). However, some longtermists also point to additional further-losses associated with transhumanism, the "argument from cosmic significance," and "ideal goods" like science, the arts, and morality (see Parfit 1984; Ord 2020). Taking these in order: many longtermists are transhumanists who believe that reengineering humanity using advanced technologies could usher in a "utopian" world of radical abundance, immortality, and superintelligence (see Bostrom 2020; Ord 2020). The future could thus be *qualitatively* better in addition to being *quantitatively* bigger. Hence, if ASI were to cause our extinction, we would lose this techno-utopian paradise that we could have otherwise created by actualizing the transhumanist project of becoming "superior" posthumans.

Our extinction would also remove the only sentient beings in the known universe who are endowed with moral and rational capacities. These capacities enable us to look up at the midnight firmament in wonder and awe, appreciate the beauty of art and nature, and act from moral reasons rather than instinct or impulse. Some further-loss theorists argue that this makes us cosmically significant, and hence that the universe would be impoverished without us. The argument from cosmic significance thus provides a second reason that some longtermists see Being Extinct as a source of badness.

With respect to the non-hedonic or "ideal" goods, there may be additional things in the world that are valuable in their own right but depend on our existence for their existence. Works of art provide an example: if humanity were to vanish, museums would gradually fall into disrepair, destroying great pieces of art that may be valuable for their own sake. To my knowledge, the first person to articulate this idea was Shelley in her aforementioned novel *The Last Man*. Lionel Verney, the protagonist, contrasts the disappearance of "man" in the collective sense with "man" in the individual sense, noting that the former would mean the concomitant loss of many valued things like knowledge, science, technology, poetry, philosophy, sculpture, painting, music, theater, laughter, and so on. "Alas!," he ex-

claims, "to enumerate the adornments of humanity, shews, *by what we have lost*, how supremely great man was. It is all over now" (Shelley 1826, italics added).

Another expression of this idea comes from Samuel Scheffler, who argues that

> there is a conservative dimension of valuing, something approaching a conceptual connection between valuing something and wanting it to be sustained and to persist over time. … This connection helps to explain part of our reaction to the prospect of humanity's imminent disappearance, for part of what is shocking about that prospect is the recognition of how much of what we value will disappear along with the human race. All of the many things we value that consist in or depend on forms of human activity will be lost when human beings become extinct. No more beautiful singing or graceful dancing or intimate friendship or warm family celebrations or hilarious jokes or gestures of kindness or displays of solidarity (Scheffler 2018).

Other further-loss theorists might also point to certain "business" being left "unfinished," e.g., constructing a complete scientific theory of the universe. Or, to quote I. F. Clarke in a 1971 article: "World peace, universal prosperity, the reign of law, the brotherhood of man—these aspirations make up the unfinished business of the human race" (Clarke 1971). The failure to achieve these ends could constitute extra losses above and beyond whatever harms Going Extinct might entail.

Still others would cite the idea of vicarious immortality, whereby one "lives on" in the minds of future people. Immortality of this sort has motivated many artists, scientists, politicians, and academics who have striven to leave a positive legacy that persists beyond their own expiration. If humanity is no more, then the memories of such people would be lost forever (see Partridge 1981). Anders takes up this idea in arguing that our extinction would cause all past people to die a "second death," such that "after this second death everything would be as if they had never been." He elaborates as follows:

> The door in front of us bears the inscription: "Nothing will have been"; and from within: "Time was an episode." Not, however, as our ancestors had hoped, an episode between two eternities; but one between two nothingnesses; between the nothingness of that which, remembered by no one, will have been as though it had never been, and the nothingness of that which will never be. And as there will be no one to tell one nothingness from the other, they will melt into one single nothingness (Anders 1961/1983).

In this passage, Anders points not just to the second death of those who have already passed, but the non-birth of those who could have otherwise been. Both are, in his view, further losses that would render our extinction very bad independent of the details of Going Extinct.

These are a few further-loss perspectives on human extinction in general, which are also applicable to the particular case of extinction caused by ASI. The key idea is that Going Extinct is only part of the story about why our extinction would be bad. Even more significant are the various losses that Being Extinct would entail, such as the loss of wellbeing, art, science, poetry, laughter, and/or the memories of those who came before us. Further-loss

theorists would thus agree with equivalence theorists that human extinction caused by an ASI catastrophe would be bad, but for a quite different set of reasons.

3.3 Pro-extinctionist views. Most pro-extinctionists would concur with equivalence and further-loss theorists that it would be very bad if Going Extinct inflicts suffering and/or cuts lives short. Many thus argue that we should avoid scenarios of Going Extinct that would involve such harms, and that bringing about our extinction in harmful ways would be morally wrong. Omnicide—a kind of killing-extinction, in Benatar's phraseology—would be impermissible. However, they differ with equivalence and further-loss theorists in claiming that the subsequent state or condition of Being Extinct would in some way be better than Being Extant, or continuing to exist. There are several mutually compatible reasons that pro-extinctionists could point at in making their case.

The first concerns philosophical pessimism, or the idea that "life is not worth living, that nothingness is better than being, or that it is worse to be than not be" (see Beiser 2016, p. 4). This was defended most famously by Arthur Schopenhauer, who contended that we are trapped in perpetual cycles of need and boredom, which produce endless suffering. There is no positive value, he claimed, and it would have been better if Earth had remained as lifeless as the moon (Schopenhauer 2017). Despite suggesting in numerous passages that human extinction would be desirable, Schopenhauer never explicitly endorsed a pro-extinctionist position (nor did he endorse antinatalism, one possible path to extinction). However, other German pessimists of the latter 19th century *were* pro-extinctionists, including the aforementioned Mainländer and his contemporary, Eduard von Hartmann. Both argued that, because existence is infused with suffering, we should try to bring about a permanent end to human life, if not all life everywhere in the universe (once this becomes possible). For Mainländer, the preferred method was celibacy plus, in some cases, suicide: "Whoever cannot endure 'the carnival hall of the world' … should leave through 'the always open door' into 'that silent night'" (Beiser 2016, p. 222).

In contrast, von Hartmann never specified a means of extinction. "Our knowledge," he wrote, "is far too imperfect, our experience too brief, and the possible analogies too defective, for us to be able, even *approximately*, to form a picture of the end of the process" (quoted in Redacted). Rather, he argued that we should continue to develop science, technology, and civilization such that, at some point in the future, we will discover an effective procedure for expunging all life in the entire universe. "Vigorously forward in the world-process as workers in the Lord's vineyard," he declared, "for it is the process alone that can bring redemption," namely, the redemption of ending the entire world-process. Indeed, since von Hartmann was an idealist, he held that the elimination of all subjectivity in the universe would cause the universe itself to cease existing, thus yielding an eternal state of what Schopenhauer memorably called the "blessed calm of nothingness" (see Redacted; Schopenhauer 2017).

The claim that existence is inherently very bad is one reason in favor of pro-extinctionism. Another concerns an empirical rather than philosophical interpretation of pessimism. This states that life and/or the world are *in fact* very bad for largely contingent reasons. Consider that every year an average of 580,000 people die violently, while 440,000 are murdered (UNODC 2023; SAS 2023). Roughly 463,000 people are raped or sexually assaulted in the US alone, and some 600,000 US children are abused each year (RAINN 2024; Seetharaman 2024). Some 840,000 children go missing annually, resulting in an average of one child disappearing every 40 seconds (NCA 2024; CCPSC 2023). Approximately 1.2 bil-

lion humans live in acute multidimensional poverty, with some 712 million suffering from extreme poverty, a figure that has risen by 23 million since 2019 (HDR 2022; WB 2024). About the same number—735 million people—are malnourished, and 25,000 people die every day from hunger or hunger-related illnesses, including 10,000 children (CW 2023; Root 2023). Two billion people don't have access to safe water, while another 150 million worldwide are homeless (UNESCO 2024; Abbas 2024). Some 1.4 billion children live on $6.85 or less per day; an estimated 50 million people are trapped in modern-day slavery; and about 1.3 million people in the US alone have survived torture, a form of suffering that, according to some survivors, has no point of reference in our normal lives (GCECP 2024; Fleck 2023; CVT 2023; Crisp 2023).

Roughly 800 million children suffer from lead poisoning each year, which causes permanent brain damage. This is about 1/3 of all children around the world (NIEHS 2024). Another 140 million people suffer from arsenic poisoning, while 18.5 million die every year from heart disease and 10 million from cancers, which amounts to some 27,600 cancer deaths every day (or 3 human beings dying per second) (CC 2024; Roser 2021). An estimated 55 million people around the world have dementia, and about 139 million are projected to have dementia by 2050 (ADI 2015). Nine million die annually from pollution; over 51 million Americans suffer from chronic pain; 50 million Americans struggle with chronic sleep disorders; and about 40 million people in the US have to take antidepressants (EC 2018; Dillinger 2023; HD 2023; Ahrnsbrak 2021). An even higher number—46.8 million—battle drug and alcohol abuse each year, with over 178,000 dying of alcohol-related diseases every 12 months (DHHS 2024). Over 258 million Americans report that "they have experienced health impacts due to stress in the prior month," while more than 91 million say that they feel so stressed-out most days that they are unable to function normally (APA 2022). Globally, 280 million people deal with depression, and 301 million suffer from anxiety disorders (Koskie 2023).

These are the statistics that empirical pessimism is based upon: the world is a waking nightmare not necessarily because there is no positive value and we are trapped in cycles of need and boredom, as Schopenhauer argued, but because things *just are* very bad. If humanity were to go extinct, all of this human suffering would disappear, which supports the pro-extinctionist tenet that Being Extinct would be better than Being Extant.[15]

Environmental considerations yield a third reason for pro-extinctionism: our systematic obliteration of the biosphere is not only imperiling our own future on Earth, but causing untold harm to billions of nonhuman organisms, ecosystems, and landscapes. If one accepts a biocentric, biospherical egalitarian, or ecocentric theory of value, then *Homo sapiens* are not the only things with intrinsic or final value. For the sake of these other things, it would be best if *Homo shiticus*—as some environmentalists call us—were to no longer exist. Though numerous environmentalists have advocated for pro-extinctionism, most are explicit that involuntary human extinction—omnicide—would *not* be morally permissible. For example, the Voluntary Human Extinction Movement (VHEMT) argues that we should stop having children until there are no more humans on Earth. Their motto is "May we live long and die out," and they do not endorse any means of eliminating our species that would cause suffering or cut lives short (VHEMT 2024). In contrast, the Gaia Liberation Front advocates for omnicide via designer pathogens.[16]

Most pro-extinctionists would thus say that if ASI were to cause our extinction through voluntary means, this would be *very good* (especially if the ASI had little or no en-

vironmental impact beyond persuading us to die out). If it were to cause our extinction through violent and/or involuntary means, then Going Extinct would be *very bad* and we should try to do whatever we can to avoid the mass slaughter of humanity. However, in the latter case, they would add that once the critical moral threshold of 100% of the human population dying has been reached, at which point Going Extinct would give way to Being Extinct, the situation would greatly improve: there would be no more human misery, nor would there be any more human-caused ecological destruction, pollution, species extinctions, and so on. That would be better, if not positively good.

4. Conclusion

The aim of this paper was to examine the question, "Would human extinction caused by an ASI be bad?" from the perspectives of the three main positions within Existential Ethics. To do this, I first outlined these three positions, and then explained how each would assess the extinction of our species if we were to create an ASI that precipitates our collective nonexistence. My hope is that this provides a helpful degree of clarity to a deceptively complex issue: nearly everyone—including pro-extinctionists—would concur that the mass murder of everyone on Earth would be extremely bad. But beyond this, opinions diverge significantly depending on which of the three main positions one accepts.

References:

Abbas, Rabeeta. 2024. "20 Countries with the Highest Homeless Population." *Yahoo! Finance*. finance.yahoo.com/news/20-countries-highest-homeless-population-180254615.html.

ADI. 2015. "Dementia Statistics." Alzheimer's Disease International. www.alzint.org/about/dementia-facts-figures/dementia-statistics/.

Ahrnsbrak, Rebecca, and Marie N. Stagnitti. 2021. "Comparison of Antidepressant and Antipsychotic Utilization and Expenditures in the U.S. Civilian Noninstitutionalized Population, 2013 and 2018." Agency for Healthcare Research and Quality. meps.ahrq.gov/data_files/publications/st534/stat534.shtml.

Altman, Sam. 2015. "Machine Intelligence, Part 1." Personal Blog. https://blog.samaltman.-com/machine-intelligence-part-1.

Anders, Günther. 1961/1983. "Commandments in the Atomic Age." In Carl Mitcham and Robert Mackey (eds.), *Philosophy and Technology: Readings in the Philosophical Problems of Technology*. New York, NY: The Free Press.

Anders, Günther. 1962. "Theses for the Atomic Age." https://www.ratical.org/ratville/JFK/Sep11PentagonsBMovie/GATftAA.pdf.

APA. 2022. “Stress in America 2022: Concerned for the Future, Beset by Inflation.” American Psychological Association. www.apa.org/news/press/releases/stress/2022/concerned-future-inflation.

Becker, Adam. 2025. *More Everything Forever: AI Overlords, Space Empires, and Silicon Valley’s Crusade to Control the Fate of Humanity*. New York: Basic Books.

Beiser, Frederick. 2016. *Weltschmerz: Pessimism in German Philosophy, 1860–1900*. Oxford, UK: Oxford University Press.

Beitrag, Ein, and Joe McCarthy. 2017. “There Could Be 2 Billion Climate Change Refugees by 2100.” *Global Citizen*. www.globalcitizen.org/de/content/2-billion-climate-change-refugees-2100/.

Benatar, David. 2006. *Better Never to Have Been: The Harm of Coming into Existence*. Oxford, UK: Oxford University Press.

Bennett, Jonathan. 1978. “On Maximizing Happiness.” In R. U. Sikora and Brian Barry (eds.), *Obligations to Future Generations*. Winwick, UK: White Horse Press.

Bostrom, Nick. 2008. “Why I Want to Be a Posthuman When I Grow Up.” In Bert Gordijn and Ruth Chadwick (eds.), *Medical Enhancement and Posthumanity*, pp. 107–36. Berlin, DE: Springer.

Bostrom, Nick. 2013. “Existential Risk Prevention as Global Priority.” *Global Policy*. 4(1): 15-31.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press.

CC. 2024. “Arsenic Poisoning.” Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/24727-arsenic-poisoning.

CCPSC.2023. “Missing and Abducted Children.” Child Crime Prevention and Safety Center. childsafety.losangelescriminallawyer.pro/missing-and-abducted-children.html.

Clarke, I.F. 1971. “The Pattern of Prediction: Forecasting: Facts and Fallibilities.” *Futures*. 3(3): 302-5.

Crisp, Roger. 2023. “Pessimism About the Future.” *Midwest Studies in Philosophy*. 46.

Curtis, Mike, and Sam Altman. 2015. “Fireside Chat - Sam Altman President, YCombinator and Mike Curtis, VP of Engineering, Airbnb.” YouTube, timestamp 8:47. https://www.youtube.com/watch?v=d6lDZpvHAoo&t=527s&ab_channel=novaclave.

CVT. 2023. “The Facts about Torture.” Center for Victims of Torture, Center for Victims of Torture. www.cvt.org/resources/facts-about-torture/.

CW. 2023. “World Hunger Facts: What You Need to Know in 2023.” Concern Worldwide US, Concern Worldwide. concernusa.org/news/world-hunger-facts/.

Dawsey, Jason. 2016. “After Hiroshima: Günther Anders and the history of anti-nuclear critique.” In Benjamin Ziemann and Matthew Grant (eds.), *Understanding the Imaginary War: Culture, Thought and Nuclear Conflict, 1945-1990*. Manchester, UK: Manchester University Press.

DHHS. 2024. “Alcohol-Related Emergencies and Deaths in the United States.” National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human Services. www.niaaa.nih.gov/alcohols-effects-health/alcohol-topics/alcohol-facts-and-statistics/alcohol-related-emergencies-and-deaths-united-states.

Dillinger, Katherine. 2023. “Chronic Pain Is Substantially More Common in the US than Diabetes, Depression and High Blood Pressure, Study Finds.” CNN. www.cnn.com/2023/05/16/health/chronic-pain-study/index.html.

EC. 2024. “Global Pollution Kills 9 Million People a Year.” European Commission. https://ec.europa.eu/newsroom/intpa/items/612355/en#:~:text=The%20deaths%20attributed%20to%20pollution,),%20cause%202.9m%20deaths.

Finneron-Burns, Elizabeth. 2017. “What’s Wrong with Human Extinction?” *Canadian Journal of Philosophy*. 47(2-3): 327-43.

Fleck, Anna. 2023. “Infographic: Countries with the Highest Prevalence of Slavery.” *Statista*. www.statista.com/chart/30666/estimated-number-of-people-in-modern-slavery-per-1000/.

Gabbatiss, Josh. 2018. “More than Quarter of World’s Land Could Become Arid Due to Global Warming, Study Says.” *The Independent*. www.independent.co.uk/climate-change/news/global-warming-world-land-arid-desertification-climate-change-study-a8139896.html.

GCECP. 2024. “Facts on Child Poverty.” Global Coalition to End Child Poverty. www.endchildhoodpoverty.org/facts-on-child-poverty.

GLF. 1994. “Statement of Purpose (A Modest Proposal),” Gaia Liberation Front. *Church of Euthanasia*. www.churchofeuthanasia.org/resources/glf/glfsop.html.

Greaves, Hilary. 2017. “Population Axiology.” *Philosophy Compass*. 12(11).

Greaves, Hilary, and William MacAskill. 2021. “The Case for Strong Longtermism.” Global Priorities Institute Working Paper. https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf.

Häggström, Olle. 2019. “Challenges to the Omohundro–Bostrom framework for AI motivations.” *foresight* 21.1: 153-166.

HD. 2023. “Over 3 Million Americans Struggle with Chronic Fatigue Syndrome.” *U.S. News and World Report*. www.usnews.com/news/health-news/articles/2023-12-11/over-3-million-americans-struggle-with-chronic-fatigue-syndrome.

HDR. 2022. “2022 Global Multidimensional Poverty Index (MPI).” Human Development Reports, United Nations. hdr.undp.org/content/2022-global-multidimensional-poverty-index-mpi#/indicies/MPI.

Jonas, Hans. 1979. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago, IL: University of Chicago Press.

Koskie, Brandi, and Crystal Raypole. 2023. “Depression Statistics: Types, Symptoms, Treatments, More.” *Healthline Media*. www.healthline.com/health/depression/facts-statistics-infographic.

Kuhlemann, Karin. 2018. “Let’s Stop Thinking We Can Tackle It When the Time Comes. We Need to Talk about Overpopulation Now.” *HuffPost UK*. www.huffingtonpost.co.uk/entry/lets-stop-thinking-we-can-tackle-it-when-the-time-comes-we-need-to-talk-about-overpopulation-now_uk_5a675db0e4b002283006fe0c.

Knutsson, Simon. 2023. “My moral view: Reducing suffering, ‘how to be’ as fundamental to morality, no positive value, cons of grand theory, and more.” Personal website. https://www.simonknutsson.com/my-moral-view/.

Kokotajlo, Daniel, Scott Alexander, Thomas Larsen, Eli Lifland, and Romeo Dean. 2025. “AI 2027.” https://ai-2027.com/.

Lavazza, Andrea, and Murilo Vilaça. 2024. “Human Extinction and AI: What We Can Learn from the Ultimate Threat.” *Philosophy & Technology*, 37(16). https://link.springer.com/article/10.1007/s13347-024-00706-2.

Luper, Steven. 2021. “Death.” In Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=death.

Mack, Katie. 2015. “Vacuum Decay: The Ultimate Catastrophe.” *Cosmos*. https://cosmosmagazine.com/science/physics/vacuum-decay-the-ultimate-catastrophe/.

McGinn, Colin. 1993. *Problems in Philosophy: The Limits of Inquiry*. Hoboken, NJ: Wiley-Blackwell.

McMahan, Jeff. 2009. “Asymmetries in the Morality of Causing People to Exist.” In Melinda Roberts and David Wasserman (eds.), *Harming Future Persons: Ethics, Genetics, and the Nonidentity Problem*. Dordrecht, NL: Springer.

Metzinger, Thomas. 2017. “Benevolent Artificial Anti-Natalism (BAAN).” Edge.org. https://www.edge.org/conversation/thomas_metzinger-benevolent-artificial-anti-natalism-baan.

Mora, Camilo, Bénédicte Dousset, et al. 2017. “Global Risk of Deadly Heat.” *Nature News*. www.nature.com/articles/nclimate3322.

Narveson, Jan. 1967. “Utilitarianism and New Generations.” *Mind*, 76(301): 62-72.

NCA. 2024. “National Child Abuse Statistics from NCA.” National Children’s Alliance, www.-nationalchildrensalliance.org/media-room/national-statistics-on-child-abuse/.

Newberry, Toby. 2021. “How Many Lives Does the Future Hold.” *Global Priorities Institute Technical Report*. https://globalprioritiesinstitute.org/wp-content/uploads/Toby- Newberry_How-many-lives-does-the-future-hold.pdf.

NIEHS. 2024. “Lead.” National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services. www.niehs.nih.gov/health/topics/agents/lead.

Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York, NY: Hachette Books.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford, UK: Oxford University Press.

Partridge, Ernest. 1981. *Responsibilities to Future Generations: Environmental Ethics*. Amherst, NY: Prometheus Books. https://books.google.de/books?id=-pNkAAAAIAAJ.

PauseAI. 2024. “List of p(doom) Values.” PauseAI website. https://pauseai.info/pdoom.

Pearce, Joshua, and Richard Parncutt. 2023. “Quantifying Global Greenhouse Gas Emissions in Human Deaths to Guide Energy Policy.” *Energies*. www.mdpi.com/1996-1073/16/16/6074.

RAINN. 2024. “Victims of Sexual Violence: Statistics.” RAINN, Rape, Abuse, and Incest National Network, www.rainn.org/statistics/victims-sexual-violence.

Root, Rebecca L. 2023. “How We Got Here: The Origins of the Global Food And ...” Devex. www.devex.com/news/how-we-got-here-the-origins-of-the-global-food-and-nutrition-crisis-105353.

Roser, Max. 2021. “Causes of Death Globally: What Do People Die From?” *Our World in Data*. ourworldindata.org/causes-of-death-treemap.

Sagan, Carl. 1983. “Nuclear War and Climatic Catastrophe: Some Policy Implications.” *Foreign Affairs*. 62(2): 257-92.

SAS. 2023. “Global Violent Deaths in 2021.” Small Arms Survey. https://www.smallarmssurvey.org/sites/default/files/SAS-GVD-2023-update-FINAL_0.pdf.

Scheffler, Samuel. 2013. *Death and the Afterlife*. Oxford, UK: Oxford University Press.

Scheffler, Samuel. 2018. *Why Worry about Future Generations*? Uehiro Series in Practical Ethics. Oxford, UK: Oxford
University Press.

Schell, Jonathan. 1982/2000. *The Fate of the Earth: And, the Abolition*. Stanford Nuclear Age Series. Stanford, CA: Stanford
University Press.

Schopenhauer, Arthur. 2017. “On the Suffering of the World.” In E. D. Klemke and Steven Cahn (eds.), *The Meaning of Life*. Oxford, UK: Oxford University Press.

Seetharaman, Deepa. 2024. “New Child Exploitation Safety Measures.” *Wall Street Journal*. https://archive.is/6bLZy.

Shelley, Mary. 1826. *The Last Man*. Henry Colburn.

Sidgwick, Henry. 1874 *The Methods of Ethics*. Donald F. Koch American Philosophy Collection. Macmillan. https://books.google.de/books?id=KVAtAAAAYAAJ.

Singer, Peter. 2021. “The hinge of history.” *Project Syndicate*. https://www.project-syndicate.org/commentary/ethical-implications-of-focusing-on-extinction-risk-by-peter-singer-2021-10

Singer, Peter, Nick Beckstead, and Matthew Wage. 2013. “Preventing Human Extinction.” Effective Altruism Forum. https://forum.effectivealtruism.org/posts/tXoE6wrEQv7GoDivb/preventing-human-extinction.

Slovic, Paul. 2007. ‘“If I look at the mass I will never act’: Psychic numbing and genocide.” *Judgment and Decision Making*. 2(2): 79-95.

Tegmark, Max. 2024. “Why We Should Build Tool AI, Not AGI.” Future of Life Institute, WebSummit 2024. https://www.youtube.com/watch?v=UWh1MIMQd1Y&t=1s&ab_channel=FutureofLifeInstitute.

Thorstad, David. 2024. Against the singularity hypothesis. *Philosophical Studies*, 1-25.

Tonn, Bruce. 2009. “Obligations to Future Generations and Acceptable Risks of Human Extinction.” *Futures*. 41(7): 427-35.

UNESCO. 2024. “Imminent Risk of a Global Water Crisis, Warns the UN World Water Development Report 2023.” UNESCO.Org, UNESCO. www.unesco.org/en/articles/imminent-risk-global-water-crisis-warns-un-world-water-development-report-2023.

UNODC. 2023. “GLOBAL STUDY ON HOMICIDE 2023.” United Nations Office on Drugs and Crime, United Nations, 2023, www.unodc.org/documents/data-and-analysis/gsh/2023/GSH23_ExSum.pdf.

Vetter, Hermann. 1968. “Discussion.” In Paul Weingartner and Gerhard Zecha (eds.), *Induction, Physics, and Ethics*. D. Reidel Publishing Company. Dordrecht, NL: D. Reidel Publishing Company.

VHEMT. 2024. Official website. https://www.vhemt.org/.

WWF. 2012. “WWF Release Groundbreaking Guide to Commodities Investing.” World Wildlife Fund. wwf.panda.org/es/?206290%2FWWF-release-groundbreaking-guide.

[1] Note that the meaning of "human extinction" is not straightforward. It could, in fact, denote a wide range of possible scenarios. I explore this important issue in Redacted.
[2] See McGinn 1993.
[3] One recent study of how artificial superintelligence could lead to catastrophe, which has received considerable attention, is Kokotajlo et al. 2025.
[4] Note that I call this the "default view" in Redacted. I now prefer the term "consensus view."
[5] This original figure was also published in Redacted.
[6] I say "enough people" because extinction through antinatalist means would not require *everyone* to stop having children. If fertility is below replacement levels, then the human population will eventually disappear. Relatedly, if the human population were to dip blow the "minimum viable population" size, which may be as low as 150 people and as high as 40,000, then there would not be enough genetic diversity for our species to persist.
[7] The Church of Euthanasia also advocates for suicide as a way of bringing about extinction (see Redacted for discussion).
[8] Other radical environmentalists have echoed this call for omnicide (see Redacted).
[9] As Knutsson writes, "I would not say that an empty world would be good," yet he also maintains that "an empty world is the best possible world" (Knutsson 2023).
[10] This shares some themes with Thomas Metzinger's "BAAN" scenario, discussed in Metzinger (2017).
[11] Note that Anders was a further-loss theorist, not an equivalence theorists. Nonetheless, he drew attention to the badness of Going Extinct.
[12] Note that not all of these individuals are equivalence theorists. I mention them because they foreground the no-ordinary-catastrophe thesis in their writings.
[13] For a discussion about what our artificial descendants might be like, and the ethics of creating artificial descendants, see Lavazza and Vilaça 2024. Note that I object to the sort of "digital eugenics"—borrowing a term from Max Tegmark (2024)—that this paper explores.
[14] Note that Singer seems to have moved away from longtermism; see Singer 2021.
[15] One reviewer of this paper helpfully noted that we should still talk about "progress" if the overall population is doing better over time, and the total number of people who are well-off is increasing (even if the total number of people who are suffering is also increasing). I think this is a valuable perspective, although my personal opinion is that there is an asymmetry between happiness and suffering such that the latter counts for more. I appreciate the feedback offered by this reviewer, though I find myself somewhat skeptical of their view, and would consider myself to be an empirical pessimist—though I could be wrong. See Redacted for reasons that I think empirical pessimism might be right.
[16] Although I am *not* a pro-extinctionist, I am somewhat sympathetic with all three arguments for this view. But I am also sympathetic with some further-loss views. Equivalence views seem to be the most correct, but my position allows for nuance in evaluating our extinction, as it draws from all three views.