Synthese (2025) 206:30
https://doi.org/10.1007/s11229-025-05094-4

ORIGINAL RESEARCH

On the extinction of humanity

Émile P. Torres[1]

Received: 12 September 2024 / Accepted: 23 May 2025

Abstract
The topic of human extinction may be of growing importance and urgency. However, much of the discussion surrounding this topic is muddled by the fact that 'human extinction' can be defined in many different ways, and indeed different conceptions of human extinction can carry quite unique implications for how one assesses the ethical and evaluative implications of our collective disappearance. There are, furthermore, several additional distinctions that (a) are crucial for such assessments, and (b) philosophers have not yet made explicit in the ethics of human extinction literature—e.g., that between the process or event of Going Extinct and the state or condition of Being Extinct. This paper outlines a framework for thinking about the ethical and evaluative aspects of human extinction, which I hope can serve as a useful foundation for future research on the topic.

Keywords Human extinction · Existential ethics · Longtermism · Transhumanism · Pro-extinctionism · Person-affecting

1 Introduction

The idea that humanity could go extinct can be traced back to ancient Greeks like Xenophanes and Empedocles, and the atomists and Stoics, though for much of Western history it was rendered unthinkable by dominant religious and philosophical views. It reemerged in the 19th century, especially following the discovery of the second law of thermodynamics, and became widely discussed after the 1954 Castle Bravo test involving thermonuclear weapons in the Marshall Islands (Torres, 2024a,

✉ Émile P. Torres
philosophytorres@gmail.com

[1] Inamori International Center for Ethics and Excellence, Case Western Reserve University, Tinkham Veale University Center, 11038 Bellflower Rd 2nd Fl, Ste 280, Cleveland, OH 44106, USA

Published online: 01 July 2025

Part I). A Google Ngram Viewer result for the keywords 'human extinction' shows that the frequency of the term began to rise in the 1950s, underwent a significant spike in the 1980s, declined once the Cold War ended, and has drastically increased since the early aughts (see Fig. 1). Never before has the idea of our extinction been more discussed, and fretted over, than right now.

Consistent with this fact, many scholars argue that the probability of human extinction this century is unprecedentedly high—perhaps orders of magnitude higher than at any period since our species emerged 300,000 years ago [1]. Nick Bostrom, a leading figure within the TESCREAL movement (Gebru and Torres 2024), estimates a 20% chance of extinction before 2100, while a 2008 survey conducted by the now-defunct Future of Humanity Institute puts the probability at 19% (Bostrom, 2005; Sandberg & Bostrom, 2008). Along similar lines, Richard Posner declared in 2004 that "human extinction is becoming a feasible scientific project," and that the near-term risk of our extinction is "significant," while the Doomsday Clock, maintained by the *Bulletin of the Atomic Scientists*, currently sits at a mere 89 seconds before midnight, or doom—the closest it has ever been since its creation in 1947 (Mecklin, 2025; Posner, 2004). Surveys of the public mirror these anxieties. One reports that 39% of Americans believe that there is at least a 50% chance of climate change causing our extinction, while another finds that 55% are "very" or "somewhat worried" that advanced AI will precipitate our collective demise (Leiserowitz et al., 2017; MUP, 2023) [2].

Consequently, public and academic debates about human extinction—its potential etiology, probability, and ethical and evaluative implications—are gaining momentum. There is, however, a major problem with these debates: few commentators are

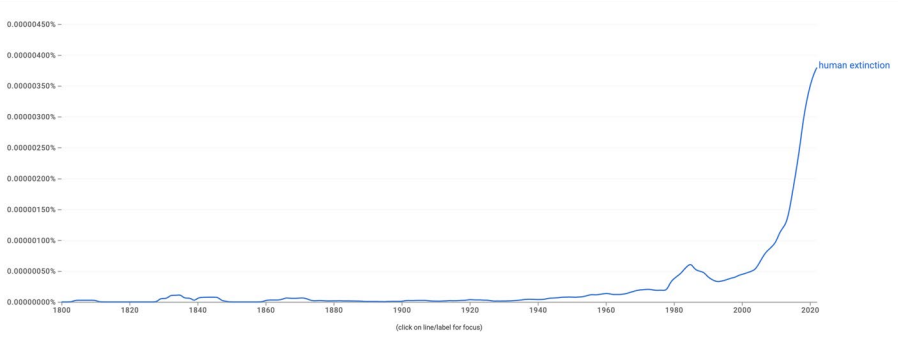

Fig. 1 Google Ngram Viewer results for "human extinction."

[1] With the possible exception of the years following the Toba catastrophe circa 75,000 years ago, when the human population may have declined to ~ 10,000 people (Gibbons, 1993).

[2] My personal view is that human extinction is extremely unlikely within the near future. I agree with Bruce Tonn's claim that "the probability of human extinction is probably fairly low, maybe one chance in tens of millions to tens of billions, given humans' abilities to adapt and survive" (Tonn, 2009b). That said, I would contend, for reasons not elaborated here, that Martin Rees' 2003 estimate that *civilization* has a 50/50 chance of collapsing this century is overly optimistic (Rees, 2003). It is very difficult to imagine how our "civilization" can survive the profusion of unprecedented threats that we'll confront over the coming decades, especially if 2 to 4 billion people die prematurely due to climate change in the next 25 years (Saye et al., 2025). But civilizational collapse need not entail human extinction.

clear about what 'human extinction' could or should mean, and most people who use the term appear to be unaware that it is polysemous: 'human' and 'extinction' could be defined in multiple ways, thus rendering the compound term doubly ambiguous [3,4].

Drawing from and elaborating upon ideas first discussed in chapter 7 of *Human Extinction: A History of the Science and Ethics of Annihilation* (2024) [5], this paper examines the various ways that 'human extinction' can be defined with respect to what I call Existential Ethics, or the study of the ethical and evaluative implications of our extinction [6]. I hope to show how certain ethical theories and normative frameworks may see one type of human extinction as being extremely bad or wrong, while simultaneously identifying other types as neutral or even desirable. Other prominent positions in Existential Ethics are indifferent about every type of human extinction except for *one*, though this is not clear from the scholarly literature. I will then introduce some additional distinctions that are crucial for making sense of the core questions of Existential Ethics, after which I will identify three major positions within the field, which I call further-loss views, equivalence views, and pro-extinctionist views. My aim is for this paper to lay a theoretical groundwork for future research on this increasingly relevant topic.

2 Disambiguating "human"

2.1 Broad versus narrow definitions

Consider that some of the philosophers who are most vocal about the importance of avoiding human extinction also endorse ethical positions that are either neutral about, or positively advocate, bringing about our extinction. Some transhumanists and longtermists, for example, would not object to the disappearance of "humanity" under certain circumstances, and indeed some actively call for the elimination of our species in the near future (see Torres, 2024b, 2025). At the same time, many of these same people argue that avoiding "human extinction" should be one of our top global priorities this century. What is going on here? Building on my initial discussion of this topic in Torres (2024a), let's begin with a look at the different ways that 'human' can be defined, and then turn to various extinction scenarios that could be instantiated given these competing definitions of 'human.'

[3] Exceptions include Elizabeth Finneron-Burns (2024), which covers similar ground as Torres (2024a), as well as Frick (2017), Fanciullo (2024), and Knutzen (2023).

[4] In this paper, I will use single quotation marks to refer to words themselves (and to indicate quotes within quotes). They can thus be read as adding 'the word(s)' before the words in single quotation marks. Example: It depends on how [the word] 'human extinction' is defined.

[5] And, more generally, Part II of that Torres (2024a).

[6] This paper should be seen as updating and superseding chapter 7 in Torres (2024a), which I now see as incomplete and partially inaccurate. For a detailed examination of the history of Existential Ethics within the Western tradition—dating back to Montesquieu's 1721 *Persian Letters*—see Part I of Torres (2024a). Some of these ideas are also explored in Torres (2023), which was written for a lay audience rather than philosophers.

The most obvious definition of 'human' or 'humanity' (I will use these interchangeably) equates it with our biological species, *Homo sapiens*[7]. 'The extinction of humanity' would thus denote "the complete disappearance of the species *Homo sapiens*." Let's call this the *Narrow Definition*. This is the definition that I believe philosophers should adopt when discussing the ethics of our extinction, for reasons explicated below.

However, many futurists prefer a more capacious definition. For example, Jason Matheny uses "'humanity' and 'humans' to mean our species and/or its descendants" (Matheny, 2007). What does 'descendants' mean? There are at least two possibilities: first, it could denote whatever beings come after us that are related to us in the right *genealogical* way. These beings might be biological, cyborgish, or wholly artificial, but so long as there is some kind of spatiotemporal continuity connecting us and them, they would constitute our descendants, the same way that we constitute the descendants of *Homo heidelbergensis*. Second, 'descendants' could refer to whatever beings come after us that are related to us in the right *causal* way. This is more general than the first interpretation, as it makes room for, e.g., populations of intelligent machines that (a) take the place of *Homo sapiens*, and (b) we don't evolve *into*, but rather create as separate, autonomous entities. In the first case, an evolutionary lineage of some sort is preserved; in the second, one might say that a new lineage is introduced alongside us, which could then supersede us. Given that Matheny is writing within the transhumanist/longtermist tradition of Existential Risk Studies, my guess is that Matheny meant to include both possibilities in his conception of 'descendants.' Hence, 'human' refers to *Homo sapiens* and whatever future populations there may be that are related to us in the right genealogical or causal ways.

Other transhumanists and longtermists stipulate definitions of 'humanity' that include an extra condition relating to moral status. For example, Nick Beckstead (2013) writes that "by 'humanity' and 'our descendants' I don't just mean the species *homo sapiens* [sic]. I mean to include any valuable successors we might have," which he later describes as "sentient beings that matter" in a moral sense. Hilary Greaves and MacAskill (2021) report 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." And Toby Ord (2020) says that "if we somehow give rise to new kinds of moral agents in the future, the term 'humanity' in my definition should be taken to include them." On these definitions, future populations of beings that *are* related to us in the right genealogical or causal ways but are *not* sentient, morally valuable, or moral agents will not count as "human." In contrast to the Narrow Definition, let's label these *Broad Definitions*.

[7] Note that there is a different sense of 'humanity' that is normative, relating to our dignity. This is relevant to the idea of normative extinction, explicated below, although I do not say much about the connection here. See the "Dead or Red?" section (pp. 293–295) of Torres (2024a) for discussion; see also Mazlish (2009) for an interesting exploration of the recent origins of our contemporary notion of "Humanity" (with a capital-'h'), as codified in international laws like those pertaining to "crimes against humanity." Clearly, there is much more to say about our notion of humanity than I have space for in this paper.

2.2 Humanity versus posthumanity

An immediate problem arises: many transhumanists and longtermists refer to future populations of our descendants that differ significantly from us as "posthumans" (see Bostrom, 2008). If *Homo sapiens* were to use radical enhancement technologies to radically reengineer our bodies and brains, then the resulting entities would belong to a new taxonomic category called "posthumanity." Yet, so long as these future beings were to instantiate the properties of being our descendants with at least comparable moral status, they would also—by stipulative definition—constitute instances of "humanity." Hence, certain future beings could simultaneously count as "human" and "posthuman," which appears to be incoherent.

This conceptual, ontological, and terminological problem has some important implications with respect to the ethics of our extinction. Consider two groups that I will call the humanists and the transhumanists. Let's say that the humanists wish to preserve our species, *Homo sapiens*, into the far future, whereas the transhumanists want to create one or more new posthuman species in the coming decades. If asked, both would affirm that avoiding human extinction is imperative. The humanists, though, will want to define 'human' as "*Homo sapiens*," since this is what matters to them, whereas the transhumanists will take it to encompass *Homo sapiens* and whatever posthuman beings we might create or become.

While it may be possible to create posthumanity without completely eliminating *Homo sapiens*, many transhumanists would not protest our species' disappearance *as long as* we were replaced by a "better" species of posthumans. As alluded to above, some explicitly argue that *Homo sapiens* should fade away for the sake of our successors. The transhumanist Steve Fuller, for example, defends "an economics of death, whereby unaugmented humans (humanity1.0) may be sacrificed for the project of creating a superior successor species," which he calls "humanity 2.0" (Thomas, 2022). Similarly, Derek Shiller (2017) argues that

> if it is within our power to provide a significantly better world for future generations at a comparatively small cost to ourselves, we have a strong moral reason to do so. One way of providing a significantly better world may involve replacing our species with something better.… Granted this assumption, it is argued that we should engineer our extinction so that our planet's resources can be devoted to making artificial creatures with better lives.

And Larry Page, the cofounder of Google, has claimed that posthuman "digital life is the natural and desirable next step in… cosmic evolution and that if we let digital minds be free rather than try to stop or enslave them, the outcome is almost certain to be good" (Tegmark 2017; see Torres, 2025, unpublished, for additional examples) [8].

Hence, some transhumanists unequivocally endorse "human extinction" while others are at least *okay with* this happening under the above-specified conditions. On the humanist view, these people should be classified as *pro-extinctionists* (or at least

[8] See Torres (2025) for other examples.

extinction neutralists; see Torres unpublished) [9]. Yet transhumanists would object, since from their perspective *Homo sapiens* could disappear entirely and forever *without* human extinction having occurred. On the Broad Definitions that they prefer, preventing human extinction does not entail that *Homo sapiens* must survive *except insofar* as our survival is necessary to create future beings that satisfy the relevant conditions of "humanity." This point is crucial because there might appear, at first glance, to be full agreement between the humanists and transhumanists about the importance of obviating "human extinction," when in fact these views differ and conflict in rather profound ways.

Here is another example: a recent survey reports that most individuals "find human extinction bad" and identify "the prevention of human extinction [to be] a key priority" (Coleman et al., unpublished). This study was coauthored by two scholars previously affiliated with the Global Priorities Institute, a longtermist/transhumanist organization at the University of Oxford. It thus seems probable that they were tacitly adopting the Broad Definitions. Yet most people, I would conjecture, intuitively adopt the Narrow Definition (see section 5). Hence, the survey's results may give the impression that most members of the public concur with the longtermists and transhumanists about preventing human extinction being a "key priority," when in fact those surveyed would likely be rather alarmed to discover that *Homo sapiens* might have no place in the longtermist/transhumanist vision of the future) [10].

This is why disambiguating 'human' and 'humanity' matters: some influential scholars use these terms in a peculiar way, which can give the false impression of convergence on central debates within Existential Ethics. If these scholars hadn't introduced an idiosyncratic interpretation of 'humanity' but had instead adopted the Narrow Definition, the present discussion would be considerably less complicated.

3 Disambiguating "extinction"

3.1 The minimal definition

There are also several distinct types of extinction scenarios that are germane to assessments of the ethical and evaluative implications of our extinction. Some of these scenarios apply to both the Narrow and Broad definitions, while others are unique to one or the other. Let's begin with a Minimal Definition of 'extinction,' and then examine various scenarios that build upon this:

[9] In Torres (unpublished), I distinguish between "extinction neutrality" and "pro-extinctionism." I argue that TESCREALists—a term that includes both transhumanists and longtermists—fall somewhere on the spectrum between extinction neutrality and pro-extinctionism. Extinction neutralists are indifferent to our extinction once posthumanity arrives; they are, I claim, pro-extinctionists *in practice*. The pro-extinctionist TESCREALists explicitly favor the extinction of our species once posthumanity arrives.

[10] As Ord (2020) writes, "forever preserving humanity as it is now may… squander our legacy, relinquishing a greater part of our potential." In other words, part of realizing our long-term potential in the universe is creating a new posthuman species that will almost certainly usurp *Homo sapiens* at some point in the future—potentially the very near future, within our lifetimes. I consider this to be pro-extinctionist *at least* in practice, as my views mostly align with the humanists rather than transhumanists.

Minimal definition: a type of thing S has gone extinct if and only if there were tokens of S at some time T1, but, at some later time T2, no such tokens exist.

This is straightforward and intuitive, and could apply to a wide range of phenomena, including languages, cultures, and biological species. It also corresponds to the standard definitions of 'extinction' found in dictionaries and science textbooks, as when Merriam-Webster defines 'extinction' as "the condition or fact of being extinct," and 'extinct' as "no longer existing" (Merriam-Webster, 2021) [11]. Similarly, Julien Delord (2007) writes that "most biologists accept the following basic definition: 'The end, the loss of existence, the disappearance of a species or the ending of a reproductive lineage.'" This leaves open the possibility of a species' disappearance being *temporary*, and indeed one of the goals of the emerging field of Resurrection Biology is to bring back into existence species that are no longer instantiated in the world. Although it is theoretically possible for humanity to be resurrected after disappearing (e.g., perhaps an extraterrestrial civilization discovers Earth and its own "resurrection biologists" bring our species back to life), this appears to be so improbable that we need not discuss it here.

3.2 Terminal and final extinction

What we need is a slightly stronger conception of human extinction that includes a condition of permanence. Let's call this *terminal extinction*, defined as having occurred if and only if humanity were to disappear *entirely and forever*[12]. This could happen on both the Narrow and Broad definitions of 'humanity.' In the former case, it would entail that there are no more instances of *Homo sapiens* in the world, and that this state of affairs is permanent. In the latter case, it would entail that there are no more instances of the class of beings that includes both *Homo sapiens* and whatever descendants we might have with comparable moral status, and that this remains the case forever. Terminal extinction is complete and permanent, though as noted the Broad (but not Narrow) Definition implies that *Homo sapiens* could vanish entirely and forever without *humanity* having disappeared. Indeed, this is precisely where humanists and transhumanists diverge: the former opposes *Homo sapiens* going ter-

[11] For an insightful discussion of different types of extinction in relation to nonhuman species, see Tanswell (2022).

[12] This is different from a weaker conception of extinction that I call *demographic extinction*, which would occur if and only if the human population were to decline until there is no one left. Demographic extinction is incompatible with phyletic extinction, since it implies that our evolutionary lineage comes to an end (rather than persisting in the form of one or more daughter species). I do not discuss demographic extinction in this paper because it is primarily of historical interest: for example, several Presocratic philosophers, such as Xenophanes and Empedocles, proposed cosmological models in which the cosmos cycles through different stages. During one of these stages, all human beings on Earth perish. At a later stage, though, we will always reemerge. Xenophanes and Empedocles thus believed that our species will someday undergo *demographic extinction* but not *terminal extinction*: our complete disappearance from the universe is always a temporary state of affairs. With respect to Resurrection Biology, the idea is that we may be able to prevent some species that have already undergone demographic extinction from undergoing terminal extinction by bringing them back into existence.

minally extinct (on the Narrow Definition), whereas the latter opposes this *unless* we produce successors with the right properties.

This brings us to a second category of extinction scenario, which I will call *final extinction*[13]. For reasons explained in subsection 3.4, let's focus solely on the Narrow Definition in the remainder of this paragraph and for the next subsection [14]. Final extinction would occur if and only if the following conditions were met: *Homo sapiens* disappears entirely and forever, *and* we do not leave behind any successors. The key idea behind this definition is that what happens after our species has disappeared could make a significant difference to how one assesses the rightness/wrongness, goodness/badness, betterness/worseness of the outcome. There are many ethical perspectives and value theories, examined in section 5, that would judge the terminal extinction of *Homo sapiens* to be bad *only if* this were to coincide with, or entail, there being no future descendants of ours. Otherwise, it would *not* be bad—which is just to say that these ethical perspectives and value theories do not identify the terminal extinction of *Homo sapiens* itself as unconditionally bad, only *conditionally* or *instrumentally* bad. The scenario they identify as unconditionally bad is final human extinction.

3.3 Examples from the literature

Consider Samuel Scheffler's (2018) argument (simplified here) that the prospect of near-term human extinction would cause many people to collapse into despair due to finding the activities that once gave value to their lives empty and meaningless. One reason is that an important source of value in our lives derives from meliorative transgenerational projects that, as such, can only be brought to fruition if there are future people who carry on these projects. The question thus becomes: must these future people be members of *Homo sapiens*? The answer appears to be "no": one can imagine posthuman beings that (a) are, as such, not members of *Homo sapiens*, (b) entirely replace *Homo sapiens*, and (c) carry on such projects. (Perhaps these future "people" would be intelligent, conscious machines that continue the work of science, further develop and appreciate the arts, etc.) Hence, on the Narrow Definition, does Scheffler's argument imply that we must avoid terminal extinction? Yes, *but only* conditionally: if *Homo sapiens* undergoes terminal extinction, this would not *in itself* undermine the various transgenerational projects that enable us to have "value-laden" lives today. What we must ultimately avoid is final extinction, since this *would* entail the destruction of these projects. Scheffler's discussion of why avoiding "human extinction" matters is thus ambiguous between these two quite distinct scenarios. Although he doesn't say it, his argument targets final rather than terminal extinction on the Narrow Definition.

[13] Note that my use of "final extinction" is different from that found in the philosophical literature on species extinctions. Many philosophers use this term to denote the "termination of a lineage" (Siipi and Finkelman 2017), which corresponds to what I call (in relation to the particular case of our species) "terminal extinction." See Wienhues et al. (2023) for a useful overview of this literature.

[14] Hence, I am shifting the terminology that I use in the next few paragraphs from "humanity" to "*Homo sapiens*," since "*Homo sapiens*" is synonymous with "humanity" on the Narrow Definition.

Another example comes from totalist utilitarianism, which is closely linked to longtermism [15]. If our sole moral obligation is to maximize welfare, and if a certain kind of posthuman species would be better able to maximize welfare than *Homo sapiens*, then totalist utilitarians should positively hope that *Homo sapiens* goes out of existence—*so long as* this coincides with the emergence of a new posthuman species that is better able to maximize welfare. In other words, the terminal extinction of *Homo sapiens* would not *itself* be bad, though the final extinction of *Homo sapiens* would be (assuming that future "people" have worthwhile lives on average). The only condition in which terminal extinction would be very bad is if it were to simultaneously instantiate, or in some other way entail, final extinction [16]. To make the relation between these scenarios explicit: final extinction entails terminal extinction, but terminal extinction need not entail final extinction.

One last example is worth mentioning: consider the position of David Benatar (2006), according to which we should bring about our extinction as soon as possible. The reason he gives is that (a) birth is always a net harm, and (b) life is overflowing with misery, though cognitive biases prevent most of us from appreciating this dismal fact. (a) does not imply that humanity should go extinct, because there could be future technologies that enable people to become "functionally immortal," and hence for humanity to persist indefinitely even in the absence of procreation (see Torres, 2020). All that (a) says is that there should be no more births, not that there should be no more people. Let us, therefore, focus on (b). If the claim is that future beings, whether posthumans or instances of *Homo sapiens*, will have lives that are sufficiently miserable, then which type of extinction should Benatar advocate for: terminal or final? Clearly, Benatar should advocate for final human extinction, as this is the only way to guarantee that there is no future human suffering [17]. Terminal extinction itself won't do the trick.

3.4 Terminal and final extinction on the broad definitions

I hope to have shown at this point that distinguishing between these two extinction scenarios is very important. However, if one accepts the Broad Definitions of 'humanity,' this distinction collapses. Final extinction references what comes after "humanity" has disappeared, but if 'humanity' is defined as "*Homo sapiens* and whatever might come after us," then terminal extinction *just is* final extinction. If "humanity" on the Broad Definition disappears entirely and forever, then there are no

[15] This is because longtermism, even in its "moderate" forms, assumes the axiology of totalism; see MacAskill (2022).

[16] Thanks to an anonymous reviewer for encouraging me to reword this sentence.

[17] The same conclusion applies to other philosophers in the pessimist tradition, such as Philipp Mainländer, Eduard von Hartmann, and Peter Wessel Zapffe, as well as to negative utilitarians and radical environmentalists (who endorse human extinction for ecological reasons, often based on biocentric, biospherical-egalitarian, or ecocentric theories of value). Terminal extinction itself, on the Narrow Definition, would not guarantee the elimination of future suffering, nor is it sufficient to ensure that the obliteration of the biosphere stops. If, for example, *Homo sapiens* disappears but we create or become a successor species that carries on our environmentally destructive activities, the problem that "extinction" is supposed to solve will persist. Hence, the target of all these perspectives is, by implication, final rather than terminal extinction.

more instances of our species *or* descendants of our species. Thus, given this definition, what matters to Scheffler, totalist utilitarians, longtermists, and Benatar is terminal extinction. For the first three, the imperative is to *avoid* terminal extinction, while for the last, it is to *bring about* terminal extinction. However, for humanists—as I am using that term here—talk of "terminal extinction" using the Broad Definition is a flawed framing that should be rejected from the outset, since what humanists care about is our particular species, *Homo sapiens*, persisting into the future. The Broad Definition directs the conversation to our species *and* successors, whereas humanists want to talk just about our species.

This highlights once again why disambiguating 'human' and 'humanity' and specifying which type of extinction one is talking about is paramount. On the Narrow Definition, the humanists will oppose terminal extinction, independent of whether it coincides with final extinction, while totalist utilitarians, longtermists, etc. will be *instrumentally* invested in avoiding terminal extinction because what they ultimately care about is avoiding final extinction. On the Broad Definition, humanists will complain that the framing is all wrong, while totalist utilitarians, longtermists, etc. will say that terminal extinction—the complete and permanent disappearance of "humanity," on this more capacious definition—is what we must strive to avert [18].

3.5 Phyletic extinction on the narrow definition

There are two additional extinction scenarios that must be identified; one pertains to the Narrow Definition and the other to the Broad Definition. The first is what I call *phyletic extinction*. This would occur if and only if *Homo sapiens* were to evolve into one or more posthuman species through a process that preserves the spatiotemporal continuity of our evolutionary lineage. As with every other type of extinction here discussed, this satisfies the Minimal Definition of 'extinction,' in the following way: at some time T1, there exist instances of *Homo sapiens* in the world but, at some later time T2, our species has evolved into a successor population that is sufficiently different for it to warrant classification as a novel species; no more instances of *Homo sapiens* remain. This contrasts with final extinction, whereby our evolutionary lineage terminates and we do not have any successors [19].

There are a few points to make about this: first, phyletic extinction foregrounds the concept of *species*, which is hotly debated among philosophers and biologists. There is no need to settle this esoteric issue here: suffice it to say that phyletic extinction would happen if future populations of our descendants were to satisfy one's preferred definition of 'species' such that they no longer count as *Homo sapiens*. Second, the case of *Homo sapiens* is completely unique within the Animal Kingdom, as we appear to be on the verge of developing technologies that could enable us to radically modify our phenotypic traits. In the long run, phyletic extinction is inevita-

[18] I borrow the phrase "at any cost" from Bostrom, who writes that, from a transhumanist/longtermist perspective, "there is one kind of catastrophe that must be avoided at any cost: *Existential risk*." One type of "existential" catastrophe is human extinction (Bostrom, 2005).

[19] It is compatible with terminal extinction, as it could be that *Homo sapiens* disappears entirely and forever after evolving into a new species.

ble due to evolutionary mechanisms like natural selection, random mutation, genetic drift, and recombination. However, in the short run, we may reengineer ourselves via some process of cyborgization, where the limit of this process would be the complete replacement of our biological substrate with an artificial one (as with "mind uploading"). Put differently, there are, in the case of *Homo sapiens*, two general possibilities for how phyletic extinction could occur: one natural and the other anthropogenic. Third, phyletic extinction is incompatible with the Broad Definitions, as alluded to above. If 'humanity' means "*Homo sapiens* plus our descendants," then it cannot be the case that "humanity" disappears by evolving into a successor species, as any such successor would also count as humanity.

Some philosophers will argue that if phyletic extinction produces a "superior" version of posthumanity, it would be positively desirable. Many transhumanists champion this view: while opposing final extinction, they would endorse certain instances of phyletic extinction [20]. Others, like the humanists, may say that, since avoiding terminal extinction is what ultimately matters, and since phyletic extinction would entail that *Homo sapiens* no longer exists, we should oppose phyletic extinction. Someone with this view might follow Johann Frick (2017) in arguing that *Homo sapiens* has final value because of its uniqueness, and that this property gives us reason to preserve our species. Perhaps the posthuman beings that take our place would also be unique, but not in the same way, and hence something would be lost even if our successors were finally valuable as well. This has a curious implication: we should prevent phyletic extinction from happening whether the cause is anthropogenic or natural. We could potentially achieve this by utilizing advanced genetic engineering methods to "fix" our genotypes within some specified range of variability, thus preserving *Homo sapiens* indefinitely.

Other humanists might hold a less dogmatic view by introducing a timescale constraint when evaluating phyletic extinction. They might argue that we should avoid terminal extinction in the near term, but if phyletic extinction were to unfold over hundreds of thousands or millions of years (thus resulting in terminal extinction a very long time from now), this would not be bad. I suspect that this view is widely held. Imagine a posthuman species that is radically different from us. If you were to ask people whether it would be good, bad, or neutral for this species to take our place through phyletic extinction 50 years from now, many would likely say that it would be bad. But if you were to ask them about this happening in piecemeal fashion across long stretches of geological time, they would probably not express any aversion to

[20] As Bostrom (2013) writes:

Above, we defined 'humanity' as Earth-originating intelligent life rather than as the particular biologically defined species *Homo sapiens*. The reason for focusing the notion of existential risk on this broader concept is that there is no reason to suppose that the biological species concept tracks what we have reason to value. If our species were to evolve, or use technology to self-modify, to such an extent that it no longer satisfied the biological criteria for species identity (such as interbreedability) with contemporary *Homo sapiens*, this need not be in any sense a catastrophe. Depending on what we changed into, such a transformation might well be very desirable. Indeed, the permanent foreclosure of any possibility of this kind of transformative change of human biological nature may itself constitute an existential catastrophe.

such an outcome [21]. Hence, one's assessment of phyletic extinction may also depend on the *temporality* of the process—an idea that we will revisit in section 4.

3.6 Normative extinction on the broad definitions

A final type of extinction concerns the Broad Definitions. Consider a scenario in which we have descendants, but they come to lack the capacity for consciousness (or sentience). Assuming that consciousness is necessary for something to possess a moral status, if one defines 'humanity' as "our species and whatever successors we might have, so long as they possess at least comparable moral status," these future beings would not count as instances of "humanity." And if these future beings do not count as "human" on the Broad Definitions, then a Minimal Definition of 'human extinction' would be satisfied. Hence, given this account of "humanity," there would be no more humans in the world, even if there existed future beings that are genealogically or causally related to us in the right ways. Let's call this *normative extinction*.

Avoiding normative extinction is of great importance from certain ethical and evaluative perspectives. Consider totalist utilitarianism once again. One of the motivations behind the Broad Definitions is the totalist axiological claim that the world is better the more total value it contains [22]. A world without "humans"—beings with moral status at least comparable to ours—would contain much less value, which would be worse, while a world with a very large number of "humans" could contain much more, which would be better. It thus matters not only that we have descendants but that these descendants count as "human." Even more, it may be important for all, or nearly all, of these "humans" to be posthumans, since posthumans could generate more value than members of our species, limited as we are by our biological constraints—i.e., their moral status could be greater than ours [23].

Of note is that the loss of value that could occur if we have descendants that don't count as "human" could be equally severe as the loss of value that would occur if we do not have any descendants at all. From this perspective, normative extinction may be comparable in badness (or wrongness) to terminal extinction, on the Broad Definitions. It follows that, for totalist utilitarians, the only two extinction scenarios

[21] This is, in fact, my own view on the matter.

[22] Note that this is not the only possible motivation. See footnote 25 for a discussion of Hans Jonas' view, which also suggests a Broad Definition of 'humanity.'

[23] Is there anything comparable to normative extinction on the Narrow Definition? Consider Elizabeth Finneron-Burns' (2024) "stupid virus" scenario. In this case, a virus spreads around the planet, "removing people's ability to reason." She also mentions that we might "become so very dependent on technology that we are robbed of our intelligence." While we "would still be *homo sapiens* [sic], [we] would no longer be capable of complex language or creating art, music, scientific discoveries, etc." In this case, we have undergone neither terminal, final, nor phyletic extinction. So, as Finneron-Burns writes, "has humanity survived in these sorts of cases?" My answer as someone who strongly prefers the Narrow Definition is, "Yes, humanity has survived, and we have a special word for this: 'dystopia.'" Hence, I would not classify the "stupid virus" scenario as "human extinction" of any sort, but rather as a dystopian outcome. In contrast, if one accepts a Broad Definition that introduces additional criteria—beyond the biological—for beings to count as "humanity," then one may very well classify this as a kind of extinction, namely, normative extinction.

that must be avoided unconditionally are normative and terminal extinction (i.e., final extinction on the Narrow Definition) [24,25].

This is why many totalist utilitarians and other further-loss theorists (see section 5) preferentially employ the Broad Definitions when talking about "human extinction." However, some of these same people also endorse phyletic extinction, as mentioned above, which thus requires them to shift conceptual frames from the Broad to the Narrow Definition. Depending on the context, such people may thus surreptitiously fluctuate between these two definitions, whereas others like the humanists tend to be consistent in their use of the Narrow Definition only. (See Fig. 2 for a diagrammatic representation of these ideas) [26].

[24] There are, of course, other ways for humanity—our species or our posthuman successors—to fail to maximize value. For example, they might live in a dystopian world in which everyone's life is less than worthwhile. The specific claim here is that terminal and normative extinction, on the Broad Definitions, are two ways—of potentially many other ways—that value could fail to be maximized. Note also that I am here assuming that normative extinction is permanent or irreversible.

[25] The exact same could be said about other further-loss views, such as that defended by Hans Jonas (1979). On Jonas' account, what matters is the perpetuation of what he calls the "moral universe" within the physical universe. The moral universe is made possible by the existence of moral agents: beings with the capacity for moral responsibility, of which we are the only known instance. Although readers of Jonas' work may assume that his view requires the continuation of *Homo sapiens*, this is not the case: the only two types of human extinction that are *sufficient* for eliminating the moral universe (assuming that we are alone in the universe) are terminal and normative extinction—or, if one adopts the Narrow Definition, final rather than terminal extinction, as the continuation of *Homo sapiens* is not itself a necessary condition for the moral universe to endure. See Torres 2024a for more.

[26] Note that there is one last type of human extinction scenario that could apply to both the Narrow and Broad definitions, which is not worth discussing in the body text of this paper: *premature extinction*. The term "premature extinction" was initially used by ecologists to describe the loss of a nonhuman species due to anthropogenic causes—i.e., before the species would have otherwise "naturally" gone extinct. So far as I am aware, the first futurist to use this term in the context of Existential Ethics (and hence in relation to *human* extinction) was Bruce Tonn in 2009; it was then foregrounded and popularized by Bostrom and Beckstead in two separate publications, both from 2013 (Beckstead, 2013; Bostrom, 2013; Tonn, 2009a). Premature extinction introduces the idea that the *timing* of human extinction, however defined, matters morally. It is not an alternative to terminal, final, and normative extinction, but rather describes a particular *way* that these types of extinction could happen. The claim would be that if, say, the terminal extinction of humanity, on the Broad Definition, were to happen *after* attaining some desired goal or completing some valued project, it may still be very bad, but it would be *less bad* than if this were to happen *before* attaining that goal or completing that project. *When* our extinction happens is important, those who endorse premature extinction would argue. For longtermists, premature extinction might occur if we undergo terminal/final extinction (depending on the definition of "humanity") prior to realizing a large fraction of "our longterm potential" in the universe. Bostrom defines it as when humanity—by which he means "Earth-originating intelligent life," an even more capacious definition than Beckstead's, Greaves and MacAskill's, and Ord's—"goes extinct… before reaching technological maturity," i.e., a state in which we achieve "capabilities affording a level of economic productivity and control over nature close to the maximum that could feasibly be achieved" (Bostrom, 2013).

Others would define premature extinction as something like terminal, final, or normative extinction that happens before "humanity" finishes certain "business," such as the business of devising a complete scientific theory of the universe (see Knutzen 2023 for discussion). In fact, the first reference to the idea of premature extinction that I am aware of comes from a 1978 paper by Jonathan Bennett, in which he suggests that we have a "prima facie obligation to ensure that important business is not left unfinished." Some have called this the "unfinished business argument" for ensuring the survival of humanity, where the most natural interpretation of "humanity" is along the lines of the Broad Definition, given that finishing much of our current business (e.g., science) does not obviously require *Homo sapiens* itself to exist. Such projects could instead be carried on by some suitable posthuman successor. In sum, the idea of premature extinction

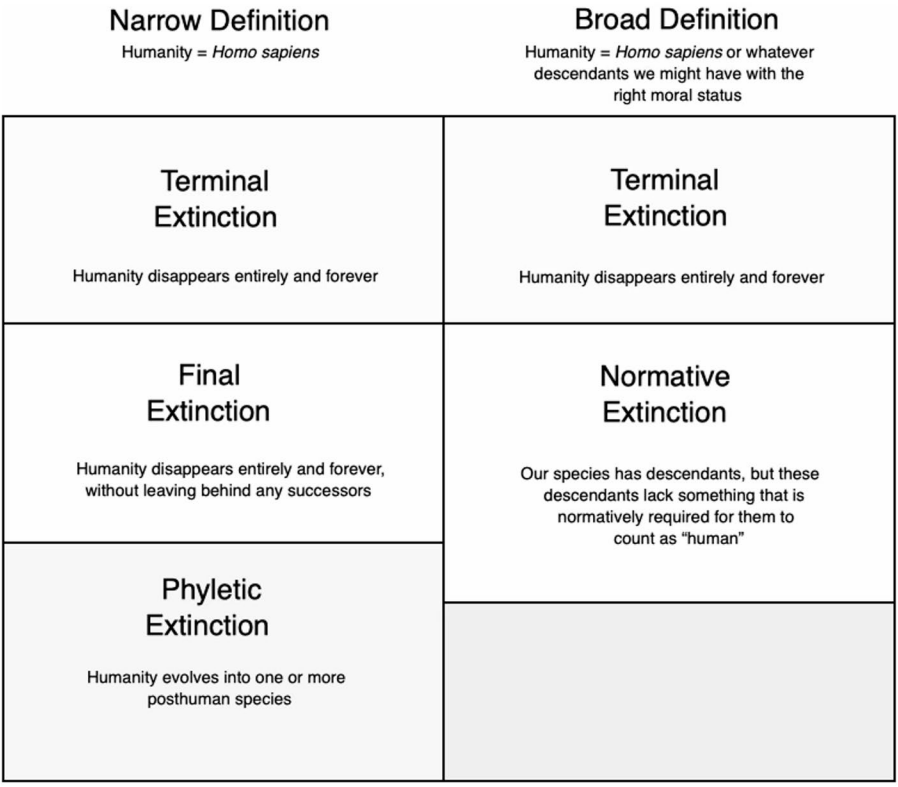

Fig. 2 Different definitions of 'humanity' combined with different extinction scenarios.

This brings us to a final set of distinctions that are crucial for navigating the labyrinth of Existential Ethics. We will then turn to the three main positions within this fledgling field.

4 Two stages of human extinction

4.1 Going versus being extinct

We have now established that the polysemy of 'human' and 'extinction' have important implications for ethical and evaluative assessments of human extinction. To my knowledge, no philosophers have made these explicit within, or noted their critical relevance to, Existential Ethics [27]. This section turns to another set of distinctions that

may be invoked whenever one accepts a vision of our future that is both normative and teleological, which is to say: a vision that identifies a future goal, or *telos*, as something that we ought to strive for. For the purposes of this paper, I will bracket the idea of premature extinction in what follows.

[27] Some of these distinctions, though, have been made within the literature on the ethics of extinction, in relation to nonhuman species. My claim is only that the distinctions made in the present article have not yet been explicitly articulated within Existential Ethics.

have been similarly neglected, but which are no less crucial for understanding the different positions that one could espouse.

The first is between (a) the process or event of Going Extinct, and (b) the subsequent state or condition of Being Extinct (analogous to the distinction between *dying* and *being dead* in the philosophical literature on death[28]). This could apply to all of the types of extinction discussed above; hence, there is Going terminally Extinct and Being terminally Extinct; Going normatively Extinct and Being normatively Extinct; and so on. As we will see in the next section, some ethical and axiological theories claim that the badness or wrongness of human extinction comes down *entirely* to the details of Going Extinct, whereas others identify Being Extinct as an *additional* source of badness/wrongness. Still other theories see most ways of Going Extinct as bad or wrong, while simultaneously claiming that Being Extinct would in some way be better than continuing to exist. One cannot make sense of the variety of positions within Existential Ethics without a clear view of the difference between Going and Being Extinct.

4.2 Three distinctions relating to going extinct

This leads to a second set of distinctions that specifically concern Going Extinct: it may matter morally *how exactly* this process or event unfolds. Consider the final extinction of *Homo sapiens* in what follows. This could result from a large asteroid striking Earth and inducing an impact winter; an engineered pandemic caused by synthetic pathogens designed in a biohacker laboratory; or an involuntary infertility scenario in which a sufficient number of people are unable to procreate. Or, it could result from everyone around the world voluntarily deciding not to have children, perhaps because they read the works of philosophical pessimists and became convinced that life is not worth perpetuating. These possibilities foreground three properties of Going Extinct (See Fig. 3).

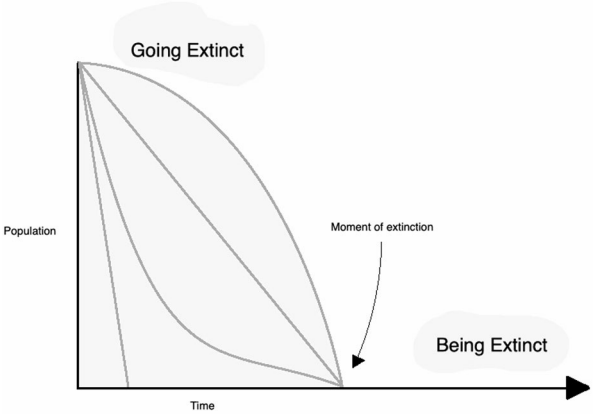

Fig. 3 Going Extinct corresponds to the human population declining, which culminates in the moment of extinction. Being Extinct is the subsequent state of nonexistence.

[28] See Luper 2024.

First, the *etiology* of our disappearance is ethically important: is the cause of our collective demise natural or anthropogenic? [29] If the former, one may still judge our extinction to be good or bad, better or worse, but ethics will have nothing to say about it. An asteroid does nothing morally wrong by crashing into Earth because it is not a moral agent, whereas someone intentionally designing a pathogen to destroy humanity does fall within the domain of ethics. This introduces additional questions, such as where to draw the line between anthropogenic and natural causes. If astronomers were to detect a 15-kilometer-wide asteroid barreling toward Earth and have the know-how to build a spacecraft to redirect this asteroid, but decide not to build the spacecraft or tell anyone else about the incoming asteroid, would the resulting extinction event be natural or anthropogenic? My sense is that it should be counted as anthropogenic, even though the hazard itself is entirely natural. But now imagine that astronomers have the means to detect asteroids, and spot one heading toward Earth. Unfortunately, they do not yet have the technology to divert this asteroid, though this technology would become feasible in five years. If scientific and technological progress had been accelerated years earlier, this technology would have been feasible when the astronomers spotted the asteroid, thus enabling them to safely divert it into space. Assuming that progress could have been accelerated but wasn't, should this scenario be considered anthropogenic or natural? I am not so sure, and I see good reasons for both possible answers.

Second, there is the question of whether our extinction is *voluntary* or *involuntary*. Some philosophers would say that final human extinction caused by a catastrophe like an engineered global pandemic would constitute a terrible moral crime, but there would be nothing bad or wrong about everyone around the world voluntarily deciding not to have children. Others would strongly disagree, arguing that *even if* our extinction were entirely voluntary, it would still be very wrong. This also introduces additional complications, such as: What conditions must be met for our extinction *to be* voluntary? If 99% of humanity were to vote in favor of extinction but 1% strongly opposed, should we count this as voluntary human extinction? Is a collective choice voluntary if a majority or plurality of people choose it? Or must it be universally chosen by literally everyone?

Yet another property of Going Extinct that may be ethically important is its *temporality*. In 1968, Hermann Vetter argued that "if mankind were extinguished by a nuclear war, the real evil… would be the way the extinction would take place: there would be so much terrible suffering for so many people before they die that this [would be] a tremendous evil," but "if mankind were completely extinguished in a millionth of a second without any suffering imposed on anybody, I should not consider this as an evil, but rather as the attainment of Nirvana" (Vetter, 1968).

On Vetter's view, *instantaneous* extinction would be positively good, while *drawn-out* extinction would be very bad, insofar as it inflicts lots of suffering on those living

[29] Note that the distinction between natural and anthropogenic extinction is central to the literature on the ethics of extinction—specifically, the extinction of nonhuman species. An important early contribution to this literature, in which the natural/anthropogenic distinction is made explicit, comes from Rolston (1985). Thanks to an anonymous reviewer for bringing my attention to this.

at the time [30]. Some philosophical pessimists who endorse human extinction would agree, as the very reason they endorse extinction is because they care about the pain, misery, and unhappiness of humanity. If Going Extinct were to impose additional suffering on people, then we should prevent ourselves from dying out in ways that entail such suffering. Negative utilitarians might disagree, arguing that an increase in human suffering, even if widespread, intense, and prolonged, might be worth it to eliminate all suffering within our future light cone [31].

The key points of this section are: assessing whether human extinction is right or wrong, good or bad, better or worse crucially depends on separating out the process or event of Going Extinct and the subsequent state or condition of Being Extinct. There are, furthermore, several distinctions to make with respect to Going Extinct, such as whether it is natural or anthropogenic, voluntary or involuntary, and instantaneous or drawn-out. As noted, this second cluster of distinctions raises many new questions, such as "What is the difference between natural and anthropogenic causes?" and "What does it mean for our extinction to be voluntary?" I will not pursue these interesting and important questions here.

5 Three positions within existential ethics

5.1 Equivalence views

Building on the previous section, we can identify three main positions within Existential Ethics. The first is what I call "equivalence views." The hallmark of these views is the claim that the badness/wrongness of "human extinction," however one defines those terms, is reducible entirely to the details of Going Extinct. If there is something bad or wrong about Going Extinct, then there is something bad or wrong about our extinction; if there is nothing bad or wrong about Going Extinct, then there is nothing bad or wrong about our extinction—full stop. Being Extinct, on these views, is morally irrelevant. Hence, if human extinction were to happen because an engineered pandemic sweeps across the globe and slowly kills everyone, then it would be very bad and/or very wrong. But if everyone around the world were to voluntarily decide not to produce another generation of children, resulting in the global population falling to zero over ~ 120 years, our extinction would be neither bad nor wrong.

Examples of equivalence views include person-affecting theories like Scanlonian contractualism and the utilitarianism of Jan Narveson (a kind of person-affecting

[30] One's position here might also depend on whether one accepts an Epicurean or anti-Epicurean view of death. If Vetter, for example, were to hold an anti-Epicurean view, according to which death can harm the one who dies, he might change his mind about instantaneous extinction. Even though no one would physically or psychologically suffer if our extinction were instantaneous and unforeseen, the resulting harm might still be enormous. This could provide a reason for objecting to instantaneous extinction.

[31] Note that instantaneous extinction is theoretically possible. For example, if the universe is in a "false vacuum" state, high-powered particle accelerators could potentially nucleate a "vacuum bubble" that expands in all directions at nearly the speed of light, destroying everything that it comes into contact with. Someone who agrees with Vetter's view might someday endorse the use of particle accelerators for this purpose.

utilitarianism, in contrast to the totalist utilitarianism of philosophers like Henry Sidgwick) [32,33]. On these accounts, Being Extinct doesn't matter because it would entail that there are no more humans in the world; and if there are no humans to be harmed or have their "interests" violated, then who has been wronged? Where is the bad? Although Going Extinct may involve grave horrors and injustices, Being Extinct plays no role in assessments of our extinction.

An intriguing implication of these views is that human extinction does not pose any unique or philosophically interesting problems. There is nothing to say about extinction *itself*—that is, everything one might wish to say about our extinction can be expressed without any reference to "extinction" at all. For example, if we undergo final extinction due to a catastrophe, this would be bad or wrong simply because catastrophes are inherently very bad or wrong (to cause). We can assess this event using our ordinary moral language about the badness/wrongness of people suffering and dying. 'Extinction' in this context is just the name we give to the limit of how bad things could get in a catastrophe, and hence 'extinction by catastrophe' conveys that the corresponding event has the highest body count possible. That may make extinction-causing catastrophes the worst type of mishap [34], but not because they result in our *extinction*—rather, because they entail the maximum number of fatalities.

To illustrate, consider the following thought experiment: in World A, there are 11 billion people; in World B, there are 10 billion people. (See Fig. 4.) An identical catastrophe happens in both worlds, killing exactly 10 billion people. At a high level of abstraction, we can ask: how many events happen in each of these worlds? In World A, a single event happens: the death of 10 billion people. In World B, two events happen: the death of 10 billion people and the extinction of humanity. We can then ask: does this extra event in World B make any ethical or evaluative difference? Is the scenario of World B *worse* than that of World A? If an omnicidal maniac named Joe were to kill 10 billion people in both worlds, would he do something *extra wrong* in World B?

Equivalence theorists will claim that the badness of the catastrophes in both worlds is the exact same. There is nothing worse about the scenario of World B compared to that of World A. Similarly, they will claim that Joe does not do something extra wrong in World B compared to World A; the wrongness of each homicidal act is also

[32] I find the current terminology in ethics to be problematic. "Total utilitarianism" is typically used as a synonym for utilitarian theories that accept a totalist axiology—which I am here calling "totalist utilitarianism." Yet Narveson, so far as I am aware, held that our moral obligation is to maximize the *total amount* of value *within the population of existing people*. This is a person-affecting version of *total* utilitarianism (see Narveson 1967). In contrast, Sidgwick held that our obligation is to maximize the *total amount* of value *within the universe as a whole*. This is an impersonalist or totalist version of total utilitarianism. The key difference concerns whether we should maximize value by creating new people: impersonalists say, "We should make people happy *and* make (new) happy people," whereas person-affecting utilitarians say, "We should focus only on making people happy."

[33] To be clear, these theories are not defined in terms of human extinction. Rather, they straightforwardly imply a particular view on the ethics of our extinction, namely, equivalence views. Thanks to an anonymous reviewer for highlighting this nuance.

[34] There are, of course, possible scenarios that are worse than catastrophic extinction—e.g., a totalitarian government takes control of the world and tortures most people for millennia. A totalitarian catastrophe might be worse than one causing final extinction.

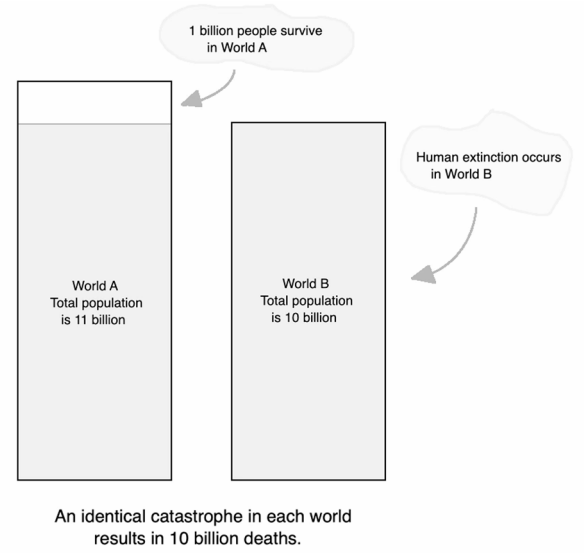

Fig. 4 The "two worlds" thought experiment.

equivalent [35]. The fact that extinction happens in World B is not important, precisely because Being Extinct is morally irrelevant. Assessing the badness/wrongness of human extinction depends entirely on the details of how it happens [36].

5.2 Further-loss views

A second class of theories in Existential Ethics is what I will call "further-loss views." These claim that Being Extinct can *also* be a source of badness/wrongness, and hence that assessing human extinction scenarios is a two-step process: first, one must examine the details of Going Extinct. Does this process or event cause physical or psychological suffering? Is it natural or anthropogenic, voluntary or involuntary, instantaneous or drawn-out? And so on. Second, one must list the various further losses associated with Being Extinct that one deems to be ethically and/or evaluatively important. Totalist utilitarians would point to all the "lost" welfare that could have otherwise existed in the future if humanity had avoided Being Extinct (in the terminal or normative senses, on the Broad Definitions). Longtermists would echo this, saying that the "opportunity costs" of Being Extinct in those ways would include the enormous number of future generations that would never be born. Some longtermists will also point to additional further losses like the "ideal goods," the transhumanist promise of a posthuman "utopia," and the apparent fact that the universe would no longer contain any rational or moral agents (Bostrom, 2020; Ord, 2020).

[35] This points to the possibility that equivalence views could have both deontic and evaluative interpretations. I will not pursue this point here.

[36] I first introduced this thought experiment in Torres (2024a). Thanks to an anonymous reviewer, I discovered while revising this manuscript that some philosophers have proposed similar thought experiments in the literature on nonhuman species extinctions. See, e.g., section 1 of Sandler (2022).

However, one does not need to be a totalist utilitarian, longtermist, or transhumanist to accept a further-loss account of extinction. Consider Hans Jonas' claim that humans, as the only known beings with the ontological capacity to be held morally responsible for their actions, are "the foothold for a moral universe in the physical world." The loss of this "universe" would constitute an immense tragedy, he says, and hence we have a moral responsibility to safeguard the continued existence of moral responsibility by ensuring humanity's survival (Jonas, 1979). This counts as a further-loss view because Jonas is contending that the badness/wrongness of extinction goes *above and beyond* whatever harms Going Extinct might involve. As with totalist utilitarianism, his view specifically targets terminal and normative extinction as what we must avoid, given a Broad Definition according to which 'humanity' might encompass any descendants of ours that possess the aforementioned ontological capacities that give rise to moral responsibility.

Another example comes from Scheffler (2018), who argues that we should prevent extinction because many of the things that we value in the world—e.g., "works of art, beautiful buildings, personal relationships… wonderful meals, beautiful concerts, thrilling conversations"—cannot exist without us (this argument is different than the one discussed earlier). In other words, their existence depends upon our existence, and hence if humanity were to disappear, so would these valued things. We have reason to prevent this from happening, Scheffler insists, because of the "conservative" nature of valuing: "What would it mean to value things but, in general, to see no reason of any kind to sustain them or retain them or preserve them or extend them into the future?" (Scheffler, 2007). This is a further-loss view because it identifies the loss of such things, entailed by Being Extinct, as an additional source of the badness/wrongness of extinction—independent of the details of Going Extinct.

So far as I know, the first explicit reference to a further-loss view in the Existential Ethics literature comes from Mary Shelley. The main character of her 1826 novel *The Last Man*, Lionel Verney, reflects at one point on the tragedy of (final) human extinction, which he is witnessing in realtime due to a natural pandemic that has spread around the world. He notes that our disappearance would mean not only that there are no humans, but that valuable phenomena like knowledge, science, technology, poetry, philosophy, sculpture, painting, music, theater, laughter, etc. would cease to be (Shelley 1826). This, he suggests, would constitute an additional tragedy on top of the misery and loneliness caused by Going Extinct.

The point to emphasize here is that further-loss views can take many forms. While totalist utilitarianism, longtermism, and transhumanism are the most influential versions today, there are many other possible interpretations of this position.

5.3 Key differences between equivalence and further-loss views

In the case of Fig. 4, further-loss theorists would argue that the scenario of World B is much worse than the scenario of World A, and that Joe does something extra wrong in World B compared to World A. *How much* worse the scenario of World B is will depend on the kind and quantity of the relevant further losses. Many further-loss theorists will argue that the losses and opportunity costs associated with Being Extinct

are profound—indeed, *far greater* than whatever harms Going Extinct might—or could possibly—involve. As Peter Singer, Nick Beckstead, and Matthew Wage write:

> 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 et al., 2013).

This suggests that human extinction would constitute a qualitatively distinct tragedy, and hence that it *does* introduce unique, philosophically important questions. An expression of this comes from Derek Parfit's (1984) thought experiment at the end of *Reasons and Persons*. He asks us to imagine three scenarios: "(1) Peace. (2) A nuclear war that kills 99% of the world's existing population. (3) A nuclear war that kills 100%." Which is the greater *difference*, he asks: between (1) and (2), or (2) and (3)? Parfit claims that most people would aver that the first difference is greatest, and indeed a recent survey finds that Parfit's hypothesis is correct (Schubert et al., 2019). Equivalence theorists would agree with these people, arguing that the difference between (2) and (3) is only 1% of humanity perishing, which is much less bad and/or wrong than 99% of people being killed. However, Parfit contends that the greater difference is between (2) and (3) because the third scenario precludes (i) the realization of all future happiness, which could be enormous, and (ii) the further development of ideal goods like the sciences, the arts, and morality. Both (i) and (ii) are further-losses that, according to Parfit, make human extinction different in kind rather than degree from non-extinction scenarios in which most, but not all, of humanity disappears.

Another way to highlight the differences between equivalence and further-loss views is to imagine a catastrophe that, over a dreadful period of 1 long year, kills more and more people until our species has undergone final extinction. (See Fig. 5.) Equivalence and further-loss theorists might agree that, as the number of fatalities increases, so does the badness of the situation. In the simplest case, they might say that twice as many deaths makes the catastrophe twice as bad.

However, something abrupt happens once the number of fatalities reaches the maximum, indicated as a "critical moral threshold" in Fig. 5. For equivalence theorists, the badness of the situation suddenly *plateaus*. This is because the moment of extinction marks the transition from Going Extinct to Being Extinct, and Being Extinct is irrelevant both ethically and evaluatively. For further-loss theorists, the badness of the situation suddenly *skyrockets*, because the moment of extinction marks the point at which all future value and opportunities are lost forever. How high these theorists will extend the vertical line depends on how large they judge the further-losses to be. If one accepts Bostrom's estimate that there could be at least 10^58 digital people in the future, one might extend the vertical line hundreds of miles above the diagram as shown, to scale, in this article [37]. Similarly, Jonas seems to consider the loss of the moral universe to be extremely bad, and hence he, too, may extend the vertical line much further than its current length in the figure; the same goes for Scheffler and Shelley. For many further-loss advocates, not only is Being Extinct morally impor-

[37] I have not actually worked out the math, but the point stands.

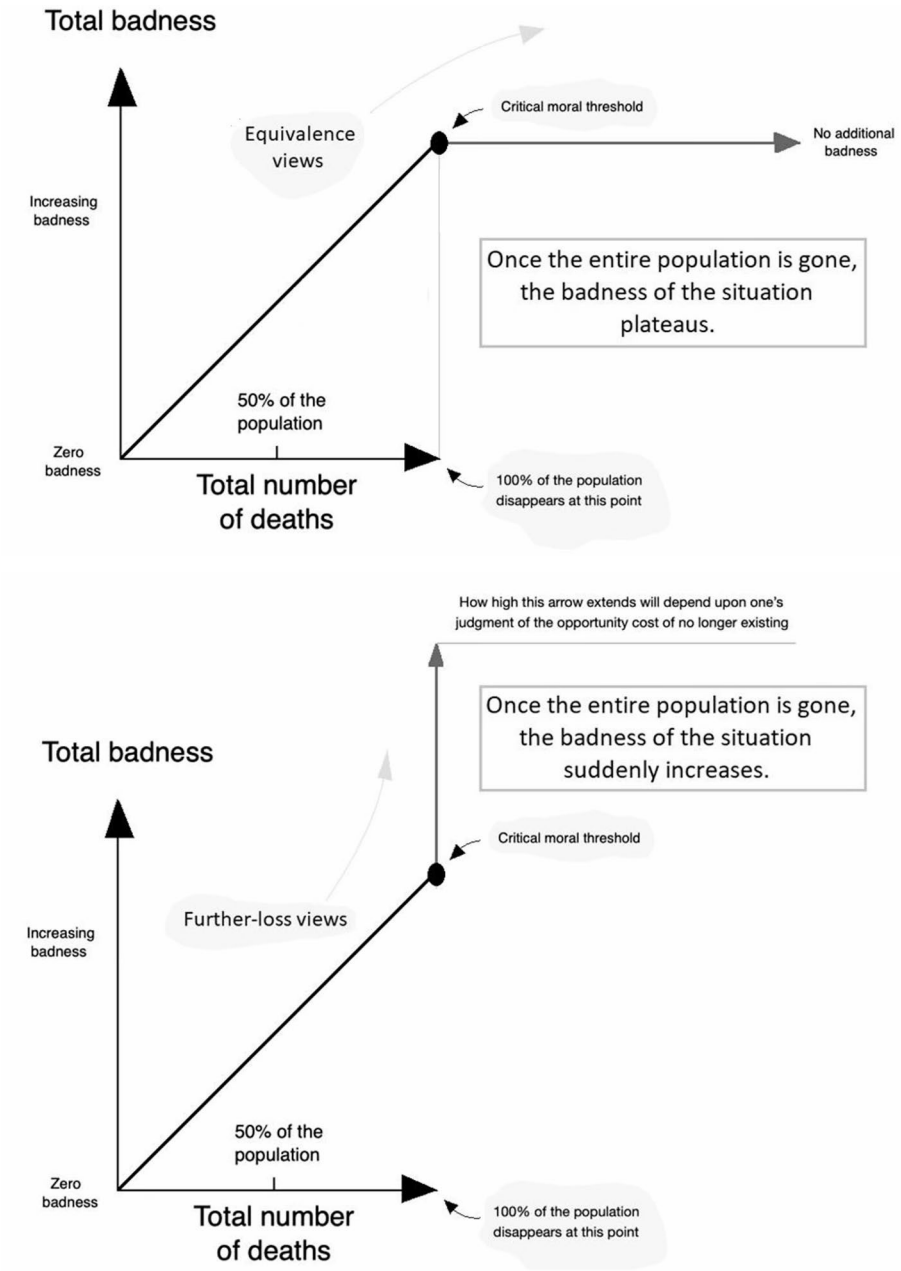

Fig. 5 Diagram showing how equivalence views (top) and further-loss views (bottom) would evaluate a catastrophic scenario in which the human population declines until no one is left.

tant, but the badness/wrongness associated with this state far surpasses the badness/wrongness of even the most atrocious ways of Going Extinct.

Yet another implication of further-loss views is that human extinction—for simplicity, let's focus again on final extinction—could be very bad or wrong *even if* there is nothing bad or wrong about Going Extinct. In Fig. 4, imagine that 10 billion people in each world voluntarily decide not to have children. Equivalence and further-loss theorists might agree that there is nothing obviously bad or wrong about this scenario in World A (though totalists could complain that the declining population would result in less total value, but let's bracket that for now [38]). However, they would vehemently disagree about the badness/wrongness of this scenario in World B, since the voluntary decision of 10 billion people not to make babies would entail our extinction. As Sidgwick (1874) declared in a discussion of celibacy, there may be nothing morally wrong about any given individual choosing to be childless, yet "a *universal refusal* to propagate the human species would be the greatest of conceivable crimes from a [totalist] Utilitarian point of view." Sidgwick would argue that voluntary childlessness in World A and World B would not be the same: the latter would constitute a profound moral crime, because unlike the former it would foreclose the possibility of all future happiness. This point about Being Extinct was later echoed by Jonathan Glover (1977) and Parfit (1984), among others.

5.4 The prototypical conception of human extinction

All of these distinctions, once again, matter greatly. Consider a question from the aforementioned survey by Coleman et al. They write that "we asked participants in both the U.S. and China whether human extinction would be good, bad, or neither" (Coleman et al., unpublished). As established in sections 2 through 4, this question is ambiguous in multiple ways: does 'human' mean "*Homo sapiens*" or "*Homo sapiens* and whatever descendants we might have with the right properties"? Furthermore, what kind of extinction are we talking about? If one adopts the Narrow Definition, 'extinction' could refer to terminal, final, or phyletic extinction; if one adopts the Broad Definition, it could denote terminal or normative extinction. Without specifying how these terms are defined, the question of Coleman et al. is hopelessly vague.

However, let's say that the question had been specifically asked about final human extinction, on the Narrow Definition. This would still be underspecified, since it also matters *how* Going Extinct takes place. My empirical hypothesis is that when most people are asked to imagine our extinction, they automatically imagine it happening due to a catastrophe. A *catastrophic etiology* is one component of what I call the "prototypical conception" of human extinction. The entire conception, I conjecture, looks something like this: "The final extinction of *Homo sapiens* brought about by a catastrophe" (where paradigmatic catastrophes include asteroid impacts and the Terminator scenario)[39].

[38] Thanks to an anonymous reviewer for pointing this out.

[39] To be clear, the prototypical conception is an empirical hypothesis. I have considerable anecdotal evidence that most people understand "human extinction" as the "final extinction of *Homo sapiens* caused

The etiological detail about catastrophic causes is important because a large majority of people will affirm that human extinction would be bad *at least* insofar as Going Extinct would induce lots of physical and/or psychological suffering [40]. At the same time, most people *also* share person-affecting intuitions that lead them to say the bigger difference in Parfit's thought experiment is between (1) and (2), rather than (2) and (3) (Schubert et al., 2019). Hence, on the assumption that human extinction is precipitated by a catastrophe, answers to Coleman et al.'s question might give the impression that there is broad agreement about the badness/wrongness of human extinction, when in fact there may be major disagreements bubbling just below the surface. For example, if the question had explicitly specified that the etiology of our extinction is that everyone voluntarily decides not to procreate, many respondents might have given a different answer, thus revealing a deeper divergence in opinions about the extinction of our species.

Here's another example to underline the point: if someone were to specifically ask me whether the final extinction of *Homo sapiens* caused by a catastrophe would be bad, I would answer: "Yes, absolutely." This is the same answer that totalist utilitarians, longtermists, and transhumanists would give, which suggests that I might be aligned with these three overlapping groups. However, I am an equivalence theorist more than anything else, and hence do not really care about *extinction itself*. In my view, the badness of the scenario in Fig. 5 plateaus once the fatalities reach the critical moral threshold of 100%. I care about people not suffering, and hence believe strongly that we must prevent our extinction *if the cause* were to be a catastrophe. But if the cause is voluntary, uncoerced, peaceful, etc., I would not object to humanity disappearing. My guess is that a large majority of people would agree with my claim about averting extinction-causing catastrophes, but far fewer would agree with the further-loss theorists' claim that voluntary extinction would constitute a horrific crime. If people were pushed to think beyond the prototypical conception, I suspect many would accept something like an equivalence view [41].

5.5 Pro-extinctionist views

The last major position within Existential Ethics is what I call "pro-extinctionism." How one understands this class of views, and which particular theories one includes within it, will depend on one's definition of 'humanity' and the extinction scenario being examined—a point revisited in the next section.

---

by a catastrophe," but this hypothesis lacks scientific evidence. I would be eager for psychologist to prove me right or wrong.

[40] In Torres (R&R), I call this the "consensus view," defining it as:

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

Previously, in Torres (2024a), I called it the "default view."

[41] Or perhaps most would assent to humanism, as I defined it, arguing that we should avoid all extinction scenarios that result in our species disappearing. This stands in opposition to equivalence views, as discussed more in section 6.

For now, let's focus on the most straightforward versions of pro-extinctionism built on ethical views like philosophical pessimism and radical environmentalism. These tend to make at least the following claim: Being Extinct would in some way, or for some reason, be *better than* Being Extant, i.e., continuing to exist (Torres unpublished). This does not mean that Being Extinct would be *good*. Simon Knutsson, for example, argues that "an empty world is the best possible world," but adds that "I would not say that an empty world would be good". A state of affairs can be better but still not-good or even quite bad. In contrast, Benatar is a pro-extinctionist who seems to believe that Being Extinct would be positively good, because it would mean the absence of both pleasure and pain, where the absence of pleasure is not bad and the absence of pain is good. This contrasts with the situation of existence, which involves the presence of both pleasure and pain, where the presence of pleasure is good and the presence of pain is bad (Benatar 2006). Not only is a good/not-bad situation (Being Extinct) better than a good/bad one (Being Extant), but it falls above the neutral line that demarcates the division between good and bad.

A key point to make about pro-extinctionism is that it is, in most cases, a claim about Being Extinct versus Being Extant, not about Going Extinct [42]. Many pro-extinctionists concur that if Going Extinct were to involve lots of suffering and/or cut lives short, it would be very bad or wrong to bring about. This is why the majority of pro-extinctionists from the German pessimists of the latter 19th century up to contemporary radical environmentalists have *opposed* omnicide, or "the murder of everyone" (see Torres, 2024a). Benatar, for instance, distinguishes between a "dying-extinction" and a "killing-extinction," where the former is, roughly speaking, voluntary whereas the latter is not. He then argues that the only morally acceptable way of bringing about our extinction—and here he appears to mean final extinction on the Narrow Definition, or terminal extinction on the Broad Definition—would be voluntary, by choosing not to have children[43].

These pro-extinctionists thus confront a practical ethical problem that is unique to their position: if Being Extinct is better than Being Extant, and if Being Extinct is something that we ought to strive for, how do we get from here to there? How should we bring about our extinction? There are three main items on the menu of options: *omnicide*, whereby one or more people kill everyone; *pro-mortalism*, whereby everyone kills themselves; and *antinatalism*, whereby everyone chooses to stop procreat-

[42] Failing to appreciate this point can lead to misleading claims, as when the journalist Dylan Matthews writes that "unless you are a member of the Voluntary Human Extinction movement, you'll probably agree that human extinction is indeed bad" (Matthews 2022). The Voluntary Human Extinction Movement (VHEMT), though, strongly opposes any form of human extinction that would be involuntary, especially if it were to entail physical or psychological suffering. Members of the community would contend that if an asteroid were barreling toward Earth, and if there were something we could do to redirect it away from Earth, we should redirect it—even though the outcome of Being Extinct would be better than Being Extant.

[43] This points to two interpretations of antinatalism: as an *ethical view* and as a *method*. One might start with pro-extinctionism and then adopt antinatalism as a method of bringing about our nonexistence. Or one might start with antinatalism as an ethical view and end up at pro-extinctionism. I gesture at this in the following paragraphs.

ing [44]. Most pro-extinctionists see antinatalism as the only permissible path to our extinction [45]. At the same time, nearly everyone admits that the antinatalist option is extremely unlikely to be implemented, as there is almost zero chance that a sufficiently large number of people around the world will voluntarily decide not to have children. Our default pronatalist tendencies are simply too entrenched. Hence, if humanity does go extinct, it will almost certainly be the result of a global catastrophe, which most pro-extinctionists agree would be very bad or wrong to bring about.

Returning once more to Fig. 4, these pro-extinctionists would say that in a forced-choice situation, the scenario of World B would be (evaluatively) preferable. While 10 billion people deciding not to make babies would be ideal, if the extinction-causing event in each world were a catastrophe that involuntarily kills 10 billion people, at least there would be a silver lining to the second scenario: humanity would no longer exist. Hence, whereas further-loss theorists will see the scenario of World B as significantly worse than that of World A, and equivalence theorists will claim that there is no difference between the two, pro-extinctionists will say that the scenario of World B is better, though most would strongly prefer that humanity dies out through some voluntary, peaceful means rather than because of a catastrophe. With respect to Fig. 5, these pro-extinctionists would say that once the catastrophe reaches the critical moral threshold of 100%, the badness of the situation neither plateaus nor skyrockets, but *declines*. That is, the transition from Being Extant to Being Extinct would mark an *improvement*, with advocates like Benatar insisting that it would be positively good.

6 Concluding discussion

This paper has delineated several distinctions and categories that are crucial for making sense of Existential Ethics. As of now, the literature is rather confusing and incoherent, with many contributors being unclear about how they define 'human' and 'humanity,' which type of extinction they are talking about, and the importance of Going Extinct versus Being Extinct.

[44] First, note that some people use the term "pro-mortalism" as a synonym of 'omnicide.' I am here distinguishing the two. Second, omnicide, pro-mortalism, and antinatalism wouldn't need to involve *everyone*, as I suggest in the body text. So long as *enough* people were involved, humanity could disappear. That is to say, if mass murder, mass suicide, or a refusal to procreate were to cause the human population to dip below the "minimum viable population," then we will have gone *functionally extinct*, at which point our eventual disappearance will be guaranteed. Note finally that one could advocate for a combination of pro-mortalism and antinatalism, whereby some people kill themselves while others decide not to procreate, a position that was endorsed by Philipp Mainländer.

[45] There are exceptions, though. As noted in the previous footnote, Philipp Mainländer also suggested a pro-mortalist route, and indeed he committed suicide days after volume I of his *magnum opus* was published. Eduard von Hartmann opposed antinatalism, instead arguing (roughly speaking) that as civilization develops, a means for eliminating all life in the universe, including all human life, will gradually come into view. Even on this view, though, Hartmann seems to have believed that total annihilation in the future ought to be voluntary: the development of our consciousness over time, he claimed, will lead a growing number of people to realize that existence is very bad, and hence that nothing should exist. Because he was an idealist, he believed that if all subjects are eliminated from the universe, the universe itself will cease to be (see Beiser 2016, ch. 7).

While most people, I believe, will intuitively understand 'human' as denoting "*Homo sapiens*," many futurists use the word to pick out a much larger class of current and future beings. The Broad Definitions tend to fit more naturally, in most contexts, with further-loss views like totalist utilitarianism and longtermism, while equivalence views may be neutral between the Narrow and Broad definitions, since all that matters to such views is whether there is anything bad or wrong about the processes or events leading up to "human extinction," however defined. Pro-extinctionist views of the sort discussed above may tend toward the Broad Definitions, since their goal is to rid the world of people who are capable of suffering or further destroying the biosphere. Meanwhile, the humanists—as I am using the term—will preferentially employ the Narrow Definition, since what they care about is the survival of our particular species.

Furthermore, different positions in Existential Ethics will identify different types of extinction as bad or wrong, while sometimes identifying other types as neutral or even good. Although I did not have space to expound this in detail, how exactly one classifies some of these positions will depend in part on which definitions of 'humanity' one uses and which extinction scenarios are under consideration [46]. As alluded to in section 3.5, many longtermists and transhumanists will see the phyletic extinction of *Homo sapiens* as a positive development, if the posthuman beings that we evolve into are better able to fulfill "our longterm potential" (Ord, 2020). On the Narrow Definition, these views should be classified as pro-extinctionist, given that some advocates explicitly argue for terminal extinction (Shiller 2017; Torres unpublished, section 4). However, on the Broad Definitions, longtermism and transhumanism constitute further-loss views that strongly oppose terminal and normative extinction, both of which they see as unconditionally bad or wrong. In either case, the nonexistence of our species itself would be neither bad nor wrong—our survival matters only insofar as it is necessary to produce a "superior" posthuman population.

With respect to the humanists, they would oppose every type of extinction on the Narrow Definition: terminal, final, and phyletic. This is not an equivalence view, since they are not claiming that the badness/wrongness of extinction comes down to the details of Going Extinct, and it is explicitly not pro-extinctionist, either. It could be interpreted as a further-loss view: as hinted in section 3.5, some might contend that our species has final value, and hence that our disappearance would constitute a tragedy above and beyond the harms of Going Extinct, rendering the universe itself impoverished. But it is not like other further-loss views that more naturally fit with the Broad Definitions; indeed, it fundamentally conflicts with further-loss positions like transhumanism and longtermism [47].

[46] Throughout this paper, I have focused on the paradigm cases of these positions. There are many nuances that I do not have space to explore—I will save these issues for a subsequent paper.

[47] One view that resists categorization within my tripartite taxonomy is Scheffler's argument from subsection 3.3. This claims that anticipatory knowledge of our extinction, in the terminal or normative senses on the Broad Definitions, could cause serious harm to people *in the present*. This is not exactly a further-loss view, because it doesn't identify some additional loss associated with Being Extinct itself, but nor is it an equivalence view, because the harms caused by anticipating our disappearance would obtain independent of how Going Extinct unfolds. It does, however, seem to be indifferent to terminal extinction on the Narrow Definition, for reasons outlined above. There is, clearly, more work to be done on this interesting topic.

Returning to the question of definitions, I would like to conclude with an exhortation for scholars to adopt, as standard practice, the Narrow Definition over the Broad Definition. The former enables more nuance in discussions about human extinction, as it allows one to distinguish between terminal, final, and phyletic extinction. It also helps clarify how ideologies like longtermism and transhumanism may constitute forms of pro-extinctionism (or at least extinction neutralism; see Torres unpublished). Though we often associate pro-extinctionism with ethical positions like philosophical pessimism and radical environmentalism, as I did above, one could argue that an especially insidious kind of pro-extinctionist fervor has become pervasive in certain corners of Silicon Valley since longtermism, transhumanism, and related TESCREAL ideologies have been embraced and promoted by influential figures (Gebru and Torres 2024; Torres, 2025, unpublished; Kirsch, 2023). By surreptitiously using the Broad Definitions, some advocates of these ideologies are able to claim that they care about avoiding "human extinction," when in fact their views are either okay with or positively endorse the disempowerment, marginalization, and eventual elimination of our species. Debates about the ethics of human extinction, in my view, would benefit if participants understood 'human' and 'humanity' as denoting *Homo sapiens*, rather than our species and whatever posthuman beings we might become or create.

This paper does not offer an exhaustive treatment of the subject [48]. However, I hope it provides a degree of conceptual clarity to the topic, which appears to be of growing interest given the climate crisis, ongoing wars involving nuclear-armed countries, and recent development of powerful AI systems.

Acknowledgements I would like to thank Gordon Schücker, Uri Eran, Simon Knutsson, and three anonymous referees for insightful feedback on earlier drafts of this paper.

Funding I received no funding for this paper.

Data availability Not applicable.

Declarations

Conflicts of interest Not applicable.

References

Beckstead, N. (2013). *On the overwhelming importance of shaping the far future*. PhD Dissertation. Rutgers the State University of New Jersey-New Brunswick. https://rucore.libraries.rutgers.edu/rutgers-lib/40469/PDF/1/play/.

Beiser, F. C. (2016). *Weltschmerz: pessimism in German philosophy, 1860–1900*. Oxford University Press.

Benatar, D. (2006). *Better never to have been: The harm of coming into existence*. Oxford University Press.

Bostrom, N. (2005). *A Philosophical quest for our biggest problems*. TED. https://www.ted.com/talks/nick_bostrom_on_our_biggest_problems.

[48] For an exploration of how the framework developed in this paper applies to a specific, concrete threat—namely, the possible annihilation of our species due to superintelligence—see Torres, R&R. The aim of this second paper is to demonstrate the usefulness of my analysis presented in this paper.

Bostrom, N. (2008). Why I want to be a posthuman when I grow up. B. Gordon & R. Chadwick (Eds.), *Medical enhancements and posthumanity*. Springer.
Bostrom, N. (2013). Existential risk prevention as global priority. *Global Policy*, *4*(1), 15–31.
Bostrom, N. (2020). Letter from Utopia. *Studies in Ethics, Law, and Technology*, *2*(1).
Bostrom, N., & Ord, T. (2006). The Reversal test: eliminating status quo bias in applied ethics. *Ethics*, *116*(4), 656–679.
Coleman, M., Caviola, L., Lewis, J., & Goodwin, G. Unpublished. *How important is the end of humanity? Lay people prioritize extinction prevention but not above all other societal issues*. https://osf.io/preprints/psyarxiv/qn7k5/download.
Delord, J. (2007). The nature of extinction. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *38*(3), 656–667.
Fanciullo, J. (2024). Why prevent human extinction?. *Philosophy and Phenomenological Research, 109*(2), 650–662.
Finneron-Burns, E. (2024). 'Humanity': Constitution, Value, and Extinction. *The Monist, 107*(2), 99–108.
Frick, J. (2017). On the survival of humanity. *Canadian Journal of Philosophy*, *47*(2–3), 344–367.
Gebru, Timnit, and Torres, Émile P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*, 29:4.
Gibbons, A. (1993). Pleistocene population explosions: a controversial method of reconstructing prehistorical populations indicates that separate modern human groups—and not a single group from africa—suddenly expanded about 50,000 years ago. *Science*, *262*(5130), 27–28.
Glover, J. (1977) Causing Death and Saving Lives. United Kingdom: Penguin.
Glover, J. (1979/1990). *Causing death and saving lives: The moral problems of abortion, infanticide, suicide, euthanasia, capital punishment, war and other life-or-death choices*. Pelican Books.
Greaves, H., & MacAskill, W. (2021). The case for strong longtermism. *Global Priorities Institute*. https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf.
Jonas, H. (1979). *The imperative of responsibility: In Search of an ethics for the technological age*. Mersion: Emergent village resources for communities of faith series. University of Chicago Press.
Kirsch, A. (2023). *The revolt against humanity: Imagining a future without us*. Columbia Global Reports.
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Rosenthal, S., & Cutler, M. (2017). *Climate change in the american mind*. Yale Program on Climate Change Communication and George Mason University Center for Climate Change Communication, https://climatecommunication.yale.edu/wp-content/uploads/2017/07/Climate-Change-American-Mind-May-2017.pdf.
Knutzen, J. (2023). Unfinished Business. *Philosophers' Imprint, 23*(1):4, 1–15.
Luper, S., "Death", *The stanford encyclopedia of philosophy* (Winter 2024 Edition), N. Z. Edward & U. Nodelman (eds.), URL=<https://plato.stanford.edu/archives/win2024/entries/death/%3E.
MacAskill, W. (2022). *What we owe the future*. Basic books.
Matheny, j. (2007). reducing the risk of human extinction. *risk analysis: An international journal*, *27*(5), 1335–1344.
Matthews, D. (2022). How effective altruism let Sam Bankman-Fried happen. *Vox*. https://www.vox.com/future-perfect/23500014/effective-altruism-sam-bankman-fried-ftx-crypto
Mazlish, B. (2009). *The idea of humanity in a global era*. Palgrave Macmillan.
Mecklin, J. (2025). 2025 Doomsday Clock Statement. *Bulletin of the Atomic Scientists*. https://thebulletin.org/doomsday-clock/2025-statement/.
Merriam-Webster. (2021). Extinct. https://www.merriam-webster.com/dictionary/extinct?utm_campaign=sd&utm_medium=serp&utm_source=jsonld.
MUP. Most expect chatgpt will be used for cheating. Monmouth University Poll. February 15, 2023. https://www.monmouth.edu/polling-institute/reports/monmouthpoll_us_021523/.
Narveson, J. (1967). Utilitarianism and new generations. *Mind*, *76*(301), 62–72.
Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.
Parfit, D. (1984). *Reasons and persons*. Oxford University Press.
Posner, R. (2004). *Catastrophe: Risk and response*. Oxford University Press.
Rees, M. (2003). *Our final hour*. Basic Books.
Rolston, H. (1985). Duties to endangered species. *BioScience, 35*(11), 718–726.
Sandberg, A., & Bostrom, N. (2008). Global catastrophic risks survey. Future of Humanity Institute, Technical Report #2008-1. https://www.fhi.ox.ac.uk/reports/2008-1.pdf.
Sandler, R. (2022). On the massness of mass extinction. *Philosophia*, *50*, 2205–2220. https://doi.org/10.1007/s11406-021-00436-1.

Saye, L., et al. (2025). Planetary solvency–finding our balance with nature. https://www.nakedcapitalism.com/wp-content/uploads/2025/01/00-planetary-solvency-finding-our-balance-with-nature-compressed.pdf.
Scheffler, S. (2007). Immigration and the significance of culture. *Philosophy & Public Affairs*, *35*(2), 93–125.
Scheffler, S. (2018). *Why worry about future generations*? *Uehiro series in practical ethics*. Oxford University Press.
Schubert, S., Cavil, L., & Faber, N. (2019). The Psychology of existential risk: Moral judgments about human extinction. *Scientific Reports*, *9*, 15100.
Shelley, M. (1826/2008). *The last man. oxford world's classics*. Oxford University Press.
Shiller, D. (2017). In defense of artificial replacement. *Bioethics, 31*(5), 393–399.
Sidgwick, H. (1874). *The Methods of ethics*. Donald F. Koch American Philosophy Collection. Macmillan. https://books.google.de/books?id=KVAtAAAAYAAJ.
Siipi, H., & Finkelman, L. (2017). The Extinction and de-extinction of species. *Philos. Technol*, *30*, 427–441. https://doi.org/10.1007/s13347-016-0244-0.
Singer, P., Beckstead, N., & Wage, M. (2013). *Preventing human extinction*. Effective Altruism Forum. https://forum.effectivealtruism.org/posts/tXoE6wrEQv7GoDivb/preventing-human-extinction.
Tanswell, F. S. (2022). The Concept of extinction: Epistemology, Responsibility, and Precaution. *Ethics, Policy & Environment*, *27*(2), 205–226. https://doi.org/10.1080/21550085.2022.2133937.
Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Vintage.
Thomas, A. (2022). The Politics and Ethics of Transhumanism: Exploring Implications for the Future in Advanced Capitalism. PhD Dissertation. University of East London. https://repository.uel.ac.uk/download/8a1d0d464e1f89ded4ace911c3470e192da96f564aaae7c68e1c787fc2ca6a23/931830/2022_PhD_Thomas.pdf.
Tonn, B. (2009a). Obligations to future generations and acceptable risks of human extinction. *Futures*, *41*(7), 427–435.
Tonn, B. (2009b). Beliefs about human extinction. *Futures*, *41*(10), 766–773.
Torres, É. (2023). The Ethics of human extinction. *Aeon*. https://aeon.co/essays/what-are-the-moral-implications-of-humanity-going-extinct.
Torres, É. (2024a). *Human extinction: A History of the science and ethics of annihilation*. Routledge.
Torres, É. (2024b). Team Human vs. Team Posthuman—Which side are you on? *Truthdig*. https://www.truthdig.com/articles/team-human-vs-team-posthuman-which-side-are-you-on/.
Torres, É. (2025). The Endgame of edgelord eschatology. *Truthdig*. https://www.truthdig.com/articles/the-endgame-of-edgelord-eschatology/.
Torres, É. R&R (revise and resubmit). "If artificial superintelligence were to cause our extinction, would that be so bad?".
Torres, É. Unpublished. Should humanity go extinct? Examining the arguments for pro-extinctionism.
Torres, P.(Émile). (2020). Can Anti-Natalists Oppose Human Extinction? The Harm-Benefit Asymmetry, Person-Uploading, and Human Enhancement. *South African Journal of Philosophy*, *39*(3), 229–245.
Vetter, H. (1968). Discussion. P. Weingartner & G. Zecha (Eds.), *Induction, Physics, and Ethics*. D. Reidel Publishing Company.
Wienhues, A., Baard, P., Donoso, A., & Oksanen, M. (2023). The ethics of species extinctions. *Cambridge Prisms: Extinction*, *1*, e23.