AMA Journal of Ethics®
August 2025, Volume 27, Number 8: E601-606

MEDICINE AND SOCIETY: PEER-REVIEWED ARTICLE

Four Key Concepts in Existential Health Care Ethics

Émile P. Torres, PhD

Abstract

Existential ethics is the study of the ethical and evaluative implications of human extinction. This article examines 4 key concepts in this emerging field: (1) Going Extinct is different from Being Extinct; (2) extinction-causing catastrophes are different in kind, not just degree, from non-extinction-causing catastrophes; (3) "human extinction" can have multiple meanings, which, when applied, can yield multiple, even conflicting, conclusions about what might constitute best future outcomes; and (4) there are historical reasons why existential ethics has tended to be ignored until recently. One goal of this article is to launch a discussion about what existential health care ethics could look like.

Key Questions and Concepts

Existential ethics is a nascent field of philosophical inquiry that focuses on the ethical and evaluative implications of human extinction. The term *existential*, as I use it, refers to "of or relating to existence," rather than to the philosophical position of existentialism. It concerns questions like the following: If we have obligations to past people that require our species to continue to survive, what is the nature and scope of these obligations? Would the loss of hypothetical future people constitute a moral tragedy? What do we mean by "human" and "extinction" in the first place?

Some experts argue that the probability of human extinction in this century is higher than at any time in the past. Estimates can range from roughly 16% to 30%.[1,2] If these estimates are accurate, it behooves scientists, philosophers, policy makers, and health professionals to take the idea of extinction more seriously than they have. This article posits that 4 key concepts from existential ethics are important to health care and health professionalism: (1) Going Extinct is different from Being Extinct; (2) extinction-causing catastrophes are different in kind, not just degree, from non-extinction-causing catastrophes; (3) "human extinction" can have multiple meanings, which, when applied, can yield multiple, even conflicting, conclusions about what might constitute best future outcomes; and (4) there are historical reasons why existential ethics has tended to be ignored until recently.

The first 2 concepts could have significant implications for health care practice and how resources within health care systems are allocated, depending on how these concepts

are interpreted. From one perspective, which has become influential over the past decade, we should prioritize health care interventions that ensure the long-term survival of humanity—millions, billions, and trillions of years into the future—which *might* entail defunding "near-termist" interventions focused on, eg, treating individuals with diseases in impoverished countries. The third concept foregrounds the important fact that there are different ways that humanity could "go extinct." One of these ways is by using biomedical advancements to radically reengineer our species to become one or more new "posthuman" species. This goal of creating posthumanity could be seen as a form of "pro-extinctionism," ie, the view that our species *should* go extinct, which has become influential within the technology sector.[3] And, finally, the fourth concept considers why existential ethics has received very little attention until quite recently. The fact that there is no established "tradition" of existential ethics—of thinking about the ethical and evaluative implications of our extinction—might seem to undermine the field's credibility: If the topic were deserving of study, then many would have already written about it. Yet, for reasons specified below, until the 19th century, almost no one in the West believed that human extinction was even *possible*.

Given that medicine will—it seems plausible to claim—play a critical role in whether our species survives this century, the ideas and insights of existential ethics research are directly relevant to how health care is, or should be, practiced. This is why, I would argue, health professionals ought to understand at least the following 4 concepts.

Going Extinct vs Being Extinct

The first concept concerns a simple but crucial distinction between the process of Going Extinct and the subsequent state of Being Extinct. Imagine asking 2 people the question: "If a pandemic were to cause our extinction, would this be bad?" Both might answer affirmatively, but their underlying reasons might differ substantially. One person might say that the *only* reason our extinction would be bad is that Going Extinct would cause horrific suffering and cut short the lives of the approximately 8 billion people who currently exist.[4] The other might say that, *in addition to* the death and suffering caused by Going Extinct, our nonexistence would prevent a potentially vast number of future people from existing. The first person would argue that the "loss" of hypothetical future people cannot be bad, since one cannot be harmed by never existing. But for the second person, the nonexistence of these future people may constitute the *worst* aspect of our extinction by far. According to Carl Sagan, if humanity survives for another 10 million years, our planet could contain 500 trillion people.[5] The loss of these people would be much worse than the deaths of roughly 8 billion individuals, however horrific that might be.

The distinction between Going Extinct and Being Extinct is thus not merely academic; it has important practical implications: If one believes that Being Extinct is *also* a source of extinction's badness—perhaps the *main* source—then one might be inclined to strongly prioritize interventions aimed at preventing extinction-causing catastrophes over those targeting non-extinction-causing catastrophes. This brings us to a second key idea in existential ethics.

Is Being Extinct a Source of Badness?

In his 1984 book, *Reasons and Persons*, Derek Parfit describes 3 scenarios: (A) peace, (B) a nuclear war that kills 99% of humanity, and (C) a nuclear war that kills 100% of humanity.[6] He then asks whether the greater difference is between (A) and (B), or between (B) and (C). Most people identify (A) and (B) as the greater difference,[7] as most

people likely focus on immediate harms that might be associated with Going Extinct, but Parfit argues that what separates (B) and (C) is enormously larger.[6] This is because (C) would preclude the realization of all future people—and hence all future value—whereas neither (A) nor (B) necessarily would. Since the amount of future value could be astronomically huge, especially if our descendants were to colonize space, (C) coincides with a fundamental discontinuity in the badness of certain catastrophe scenarios. As the number of deaths caused by the nuclear war increases, so does the badness of the situation. However, once the last remaining human perishes, the badness of the situation suddenly skyrockets, because it is at *this particular moment* that all future value is lost forever.

In contrast to the second person's emphasis on lost future value, the first person above, who believes that the badness of human extinction is reducible entirely to the details of Going Extinct, would argue that once the last remaining human perishes, the badness of the situation *plateaus*. This is because they—contra Parfit—do not believe that the "lost" people and value associated with Being Extinct are morally relevant. As alluded to above, some who hold this view would argue that if there is no one around to suffer the nonexistence of humanity, then no one can be harmed by Being Extinct. And if Being Extinct harms no one, then Being Extinct itself cannot be bad (or wrong). This is a fundamental disagreement within existential ethics.

The connection with health care is that if there *is* a fundamental discontinuity between extinction-causing and non-extinction-causing catastrophes (ie, Being Extinct is a source of badness), then we should allocate more resources to prevent the former, even if this means neglecting the latter. Hence, health professionals who accept Parfit's view should deprioritize non-extinction-related interventions that use up valuable resources if those resources could be utilized instead to safeguard humanity's long-term survival. This reasoning is why the "longtermist" Nick Beckstead, who agrees with Parfit's view, argues that we should prioritize saving the lives of people in rich countries over saving the lives of people in poor countries, all other things being equal, given that (a) people in rich countries are better positioned to protect our long-term future and (b) the long-term future is of "overwhelming" moral importance.[8]

This conclusion follows even if the probability of any particular extinction scenario is minuscule, as the associated "existential risk" may still be very large. That is to say, a low-probability event that could result in the loss of enormous amounts of future value would still count as very "risky" on the standard definition of risk as the probability of an event multiplied by its consequences.[9] If Parfit and Beckstead are correct, then much of the current focus in health care—indeed, of our philanthropic efforts more generally—might be misguided. This is not to say that near-term individual care doesn't matter, but that it only matters insofar as it advances the aim of fulfilling our long-term "potential" in the universe over the coming millions, billions, and trillions of years—which, of course, requires that humanity does not go extinct anytime soon.

How Should *Human Extinction* Be Defined?

The third key idea concerns various ways that *human extinction* could be defined. Most people understand *human extinction* as denoting a situation in which our species, *Homo sapiens*, disappears entirely and forever. However, many futurists—including those sympathetic to Parfit's discontinuity thesis—define *humanity* as including not just *Homo sapiens* but any successors we might have, even if these beings are very different from us—eg, are entirely nonbiological in nature. Some add that our successors must also

possess certain properties to count as human, such as having a "moral status" comparable to ours.[10]

This broader definition implies that *Homo sapiens* could disappear entirely and forever, perhaps in the near future, without human extinction having occurred. As long as our disappearance coincides with the emergence of a new successor species, then "humanity" will persist. Consequently, people who define *human* or *humanity* differently might appear to agree about the importance of avoiding human extinction, yet their views could be diametrically opposed. One person might wish to preserve our particular species, while another might be indifferent to the survival of our species or even favor the active replacement of *Homo sapiens* with a successor species that they consider to be "superior."[11]

In fact, many of the loudest voices calling for efforts to avoid human extinction are "transhumanists" and longtermists who believe that creating a new posthuman species is integral to fulfilling our long-term cosmic potential.[1] (The terminology here is confusing, since posthumans would also count as humans on their definition.) Once posthumanity arrives, these transhumanists and longtermists are largely indifferent to the fate of *Homo sapiens*, and, indeed, some explicitly argue that our species *should* die out.[11,12] This is not a fringe view: The idea that the future of humanity is digital rather than biological is widespread among many in the technology sector, some of whom contend that replacing our species in the near future with artificial beings is "the natural and desirable next step in … cosmic evolution."[13]

If one understands *humanity* as referring specifically to our species, then such views, embraced by leading transhumanists and longtermists, should be categorized as pro-extinctionist, given that *Homo sapiens* would likely not survive a world run and ruled by our posthumans successors. While many people associate pro-extinctionism with positions like philosophical pessimism (nonexistence is preferable to existence; life is not worth living) and radical environmentalism (*Homo sapiens* should die out because we are destroying the biosphere), there is a particularly insidious form of pro-extinctionism that has become pervasive within powerful corners of Big Tech associated with transhumanist and longtermist ideologies.

Disambiguating the term *human extinction* is thus crucial for making sense of contemporary debates about the topic. The statement, "I oppose human extinction," is meaningless without further details about what one means by this term, as well as which aspect of extinction—Going Extinct or Being Extinct—one identifies as morally important sources of the badness or wrongness of our extinction.

Why Has Existential Ethics Been Neglected Until Quite Recently?

The fourth key idea concerns the historical question of why existential ethics has been largely neglected by academics until quite recently. For much of Western history, most people would have claimed that human extinction is fundamentally impossible. This was the case for 2 main reasons: First, most people accepted a model of reality called the "Great Chain of Being," which denies the possibility of *any kind* of extinction. The eternal completeness of the Great Chain was taken as reflecting God's perfection, and, hence, since God is perfect, no link in the chain could ever go missing. If extinction in general is impossible, then so is the extinction of humanity. This idea was immensely influential from the early first millennium until the early 1800s, when Georges Cuvier and others demolished it.[12] Second, most people throughout Western history also accepted a

Christian worldview according to which human extinction isn't part of God's plan for humanity. Our world will eventually end, but this ending will mark a glorious new beginning: eternal life in heaven for believers. Humanity cannot simply disappear entirely and forever. It wasn't until the 19th century, shortly after the idea of the Great Chain collapsed, that Christianity began to decline among the educated classes. This decline opened up conceptual space for people to accept what had, up to that point, been unthinkable: that human extinction is possible.[12]

For these 2 reasons, questions about the ethical and evaluative implications of human extinction were largely ignored within the Western intellectual tradition: What is the point of examining the ethics of something that cannot happen? However, over the past 20 years, existential ethics has begun to coalesce into a coherent field of inquiry. Much of the existential ethics discussion has been dominated by transhumanists and longtermists who accept Parfit's thesis that the badness of a catastrophe skyrockets the moment that 100% of humanity dies out—precisely because Being Extinct would prevent us from fulfilling our long-term potential in the universe.

Longtermism and transhumanism are just 2 of many positions one could espouse within existential ethics. Only since 2015 or so have philosophers begun to systematically explore a range of alternative views. This topic's novelty means there is no well-established, time-honored tradition of thinking about it rigorously. This is unfortunate because there are plausible arguments for the claim that human extinction may be more probable this century than ever before in our history, given the novel threats posed by thermonuclear war, engineered pandemics, and perhaps artificial superintelligence.

Conclusion

In this short article, I have outlined 4 key concepts in existential ethics. My hope is that this exposition provides a useful point of departure for future discussions about the important, yet underexplored, connections between health care and the ethical and evaluative implications of human extinction. If our species really could disappear this century—as some advocates of ideologies like transhumanism, longtermism, and "accelerationism" hope will happen—then surely it behooves philosophers and health professionals alike to examine the nature and implications of this event.

References

1. Ord T. *The Precipice: Existential Risk and the Future of Humanity*. Hachette Books; 2020.
2. Karger E, Rosenberg J, Jacobs Z, Hickman M, Tetlock P. Subjective-probability forecasts of existential risk: initial results from a hybrid persuasion-forecasting tournament. *Int J Forecast*. 2025;41(2):499-516.
3. Gebru T, Torres ÉP. The TESCREAL bundle: eugenics and the promise of utopia through artificial general intelligence. *First Monday*. 2024;29(4):13636.
4. Current world population. Worldometer. Accessed on March 14, 2025. https://www.worldometers.info/world-population/
5. Sagan C. Nuclear war and climatic catastrophe: some policy implications. *Foreign Affairs*. December 1, 1983. Accessed February 3, 2025. https://www.foreignaffairs.com/articles/1983-12-01/nuclear-war-and-climatic-catastrophe-some-policy-implications
6. Parfit D. *Reasons and Persons*. Oxford University Press; 1987.
7. Schubert S, Caviola L, Faber NS. The psychology of existential risk: moral judgments about human extinction. *Sci Rep*. 2019;9(1):15100.

8. Beckstead N. *On the Overwhelming Importance of Shaping the Far Future*. Dissertation. Rutgers University; 2013. Accessed February 3, 2025. https://rucore.libraries.rutgers.edu/rutgers-lib/40469/
9. Boyd M, Wilson N. Assumptions, uncertainty, and catastrophic/existential risk: national risk assessments need improved methods and stakeholder engagement. *Risk Anal*. 2023;43(12):2486-2502.
10. Hilary G, MacAskill W. The case for strong longtermism. Global Priorities Institute, University of Oxford; 2021. GPI working paper 5-2021. Accessed February 3, 2025. https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf
11. Shiller D. In defense of artificial replacement. *Bioethics*. 2017;31(5):393-399.
12. Torres ÉP. *Human Extinction: A History of the Science and Ethics of Annihilation*. Routledge/Taylor & Francis; 2023.
13. Tegmark M. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf Doubleday; 2017.

Émile P. Torres, PhD is a postdoctoral scholar at the Inamori International Center for Ethics and Excellence at Case Western Reserve University in Cleveland, Ohio. Their work focuses on existential threats to humanity and civilization, as well as on the ethical and evaluative implications of human extinction. They have published widely in both academic journals and popular media.

Citation
AMA J Ethics. 2025;27(8):E601-606.

DOI
10.1001/amajethics.2025.601.

Conflict of Interest Disclosure
Contributor disclosed no conflicts of interest relevant to the content.

ISSN 2376-6980