# Astronomical Waste

## The Opportunity Cost of Delayed Technological Development

Nick Bostrom
Oxford University
www.nickbostrom.com

**Abstract**

With very advanced technology, a very large population of people living happy lives could be sustained in the accessible region of the universe. For every year that development of such technologies and colonization of the universe is delayed, there is therefore a corresponding opportunity cost: a potential good, lives worth living, is not being realized. Given some plausible assumptions, this cost is extremely large. However, the lesson for standard utilitarians is not that we ought to maximize the pace of technological development, but rather that we ought to maximize its safety, i.e. the probability that colonization will eventually occur. This goal has such high utility that standard utilitarians ought to focus *all* their efforts on it. Utilitarians of a "person-affecting" stripe should accept a modified version of this conclusion. Some mixed ethical views, which combine utilitarian considerations with other criteria, will also be committed to a similar bottom line.

Originally published in *Utilitas* Vol. 15, No. 3 (2003): pp. 308-314

Translations: Bosnian, French by Neil Warren, German, Portuguese, Russian, Spanish
Formats: HTML, PDF, Audio

---

## I. The Rate of Loss of Potential Lives

As I write these words, suns are illuminating and heating empty rooms, unused energy is being flushed down black holes, and our great common endowment of negentropy is being irreversibly degraded into entropy on a cosmic scale. These are resources that an advanced civilization could have used to create value-structures, such as sentient beings living worthwhile lives.

The rate of this loss boggles the mind. One recent paper speculates, using loose theoretical considerations based on the rate of increase of entropy, that the loss of potential human lives in our own galactic supercluster is at least $\sim 10^{46}$ per

century of delayed colonization.[1] This estimate assumes that all the lost entropy could have been used for productive purposes, although no currently known technological mechanisms are even remotely capable of doing that. Since the estimate is meant to be a lower bound, this radically unconservative assumption is undesirable.

We can, however, get a lower bound more straightforwardly by simply counting the number or stars in our galactic supercluster and multiplying this number with the amount of computing power that the resources of each star could be used to generate using technologies for whose feasibility a strong case has already been made. We can then divide this total with the estimated amount of computing power needed to simulate one human life.

As a rough approximation, let us say the Virgo Supercluster contains $10^{13}$ stars. One estimate of the computing power extractable from a star and with an associated planet-sized computational structure, using advanced molecular nanotechnology[2], is $10^{42}$ operations per second.[3] A typical estimate of the human brain's processing power is roughly $10^{17}$ operations per second or less.[4] Not much more seems to be needed to simulate the relevant parts of the environment in sufficient detail to enable the simulated minds to have experiences indistinguishable from typical current human experiences.[5] Given these estimates, it follows that the potential for approximately $10^{38}$ human lives is lost every century that colonization of our local supercluster is delayed; or equivalently, about $10^{29}$ potential human lives per second.

While this estimate is conservative in that it assumes only computational mechanisms whose implementation has been at least outlined in the literature, it is useful to have an even more conservative estimate that does not assume a non-biological instantiation of the potential persons. Suppose that about $10^{10}$ biological humans could be sustained around an average star. Then the Virgo Supercluster could contain $10^{23}$ biological humans. This corresponds to a loss of potential of over $10^{13}$ potential human lives per second of delayed colonization.

What matters for present purposes is not the exact numbers but the fact that they are huge. Even with the most conservative estimate, assuming a biological implementation of all persons, the potential for over ten trillion potential human beings is lost for every second of postponement of colonization of our

[1] M. Cirkovic, "Cosmological Forecast and its Practical Significance", *Journal of Evolution and Technology*, xii (2002), https://www.jetpress.org/volume12/CosmologicalForecast.pdf.

[2] K. E. Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, New York, John Wiley & Sons, Inc., 1992.

[3] R. J. Bradbury, "Matrioshka Brains", *Manuscript*, 2002, http://www.aeiveos.com/~bradbury/MatrioshkaBrains/MatrioshkaBrains.html

[4] N. Bostrom, "How Long Before Superintelligence?", *International Journal of Futures Studies*, ii (1998); R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York, Viking, 1999. The lower estimate is in H. Moravec, *Robot: Mere Machine to Transcendent Mind*, Oxford, 1999.

[5] N. Bostrom, "Are You Living in a Simulation?", *Philosophical Quarterly*, liii (211). See also https://simulation-argument.com.

supercluster.[6]

## II. The Opportunity Cost of Delayed Colonization

From a utilitarian perspective, this huge loss of potential human lives constitutes a correspondingly huge loss of potential value. I am assuming here that the human lives that could have been created would have been worthwhile ones. Since it is commonly supposed that even current human lives are typically worthwhile, this is a weak assumption. Any civilization advanced enough to colonize the local supercluster would likely also have the ability to establish at least the minimally favorable conditions required for future lives to be worth living.

The effect on total value, then, seems greater for actions that accelerate technological development than for practically any other possible action. Advancing technology (or its enabling factors, such as economic productivity) even by such a tiny amount that it leads to colonization of the local supercluster just *one second* earlier than would otherwise have happened amounts to bringing about more than $10^{29}$ human lives (or $10^{13}$ human lives if we use the most conservative lower bound) that would not otherwise have existed. Few other philanthropic causes could hope to match that level of utilitarian payoff.

Utilitarians are not the only ones who should strongly oppose astronomical waste. There are many views about what has value that would concur with the assessment that the current rate of wastage constitutes an enormous loss of potential value. For example, we can take a thicker conception of human welfare than commonly supposed by utilitarians (whether of a hedonistic, experientialist, or desire-satisfactionist bent), such as a conception that locates value also in human flourishing, meaningful relationships, noble character, individual expression, aesthetic appreciation, and so forth. So long as the evaluation function is aggregative (does not count one person's welfare for less just because there are many other persons in existence who also enjoy happy lives) and is not relativized to a particular point in time (no time-discounting), the conclusion will hold.

These conditions can be relaxed further. Even if the welfare function is not perfectly aggregative (perhaps because one component of the good is diversity, the marginal rate of production of which might decline with increasing population size), it can still yield a similar bottom line provided only that at least some significant component of the good is sufficiently aggregative. Similarly, some degree of time-discounting future goods could be accommodated without changing the conclusion.[7]

[6] The Virgo Supercluster contains only a small part of the colonizable resources in the universe, but it is sufficiently big to make the point. The bigger the region we consider, the less certain we can be that significant parts of it will not have been colonized by a civilization of non-terrestrial origin by the time we could get there.

[7] Utilitarians commonly regard time-discounting as inappropriate in evaluating moral goods (see e.g. R. B. Brandt, *Morality, Utilitarianism, and Rights*, Cambridge, 1992, pp. 23f.). However, it is not clear that utilitarians can avoid compromising on this principle in view of

## III. The Chief Goal for Utilitarians Should Be to Reduce Existential Risk

In light of the above discussion, it may seem as if a utilitarian ought to focus her efforts on accelerating technological development. The payoff from even a very slight success in this endeavor is so enormous that it dwarfs that of almost any other activity. We appear to have a utilitarian argument for the greatest possible urgency of technological development.

However, the true lesson is a different one. If what we are concerned with is (something like) maximizing the expected number of worthwhile lives that we will create, then in addition to the opportunity cost of delayed colonization, we have to take into account the risk of failure to colonize at all. We might fall victim to an *existential risk*, one where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.[8] Because the lifespan of galaxies is measured in billions of years, whereas the time-scale of any delays that we could realistically affect would rather be measured in years or decades, the consideration of risk trumps the consideration of opportunity cost. For example, a single percentage point of reduction of existential risks would be worth (from a utilitarian expected utility point-of-view) a delay of over 10 million years.

Therefore, if our actions have even the slightest effect on *the probability* of eventual colonization, this will outweigh their effect on *when* colonization takes place. For standard utilitarians, priority number one, two, three and four should consequently be to reduce existential risk. The utilitarian imperative "Maximize expected aggregate utility!" can be simplified to the maxim "Minimize existential risk!".

## IV. Implications for Aggregative Person-Affecting Views

The argument above presupposes that our concern is to maximize the total amount of well-being. Suppose instead that we adopt a "person-affecting" version of utilitarianism, according to which our obligations are primarily towards currently existing persons and to those persons that will come to exist.[9] On such a person-affecting view, human extinction would be bad only because it makes past or ongoing lives worse, not because it constitutes a loss of potential worthwhile lives. What ought someone who embraces this doctrine do? Should he emphasize speed or safety, or something else?

To answer this, we need to consider some further matters. Suppose one thinks

the possibility that our actions could conceivably have consequences for an infinite number of persons (a possibility that we set aside for the purposes of this paper).

[8] N. Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards", *Journal of Evolution and Technology*, ix (2002), https://www.jetpress.org/volume9/risks.html.

[9] This formulation of the position is not necessarily the best possible one, but it is simple and will serve for the purposes of this paper.

that the probability is negligible that any existing person will survive long enough to get to use a significant portion of the accessible astronomical resources, which, as described in opening section of this paper, are gradually going to waste. Then one's reason for minimizing existential risk is that sudden extinction would off cut an average of, say, 40 years from each of the current (six billion or so) human lives.[10] While this would certainly be a large disaster, it is in the same big ballpark as other ongoing human tragedies, such as world poverty, hunger and disease. On this assumption, then, a person-affecting utilitarian should regard reducing existential risk as a very important but not completely dominating concern. There would in this case be no easy answer to what he ought to do. Where he ought to focus his efforts would depend on detailed calculations about which area of philanthropic activity he would happen to be best placed to make a contribution to.

Arguably, however, we ought to assign a non-negligible probability to some current people surviving long enough to reap the benefits of a cosmic diaspora. A so-called technological "singularity" might occur in our natural lifetime[11], or there could be a breakthrough in life-extension, brought about, perhaps, as result of machine-phase nanotechnology that would give us unprecedented control over the biochemical processes in our bodies and enable us to halt and reverse the aging process.[12] Many leading technologists and futurist thinkers give a fairly high probability to these developments happening within the next several decades.[13] Even if you are skeptical about their prognostications, you should consider the poor track record of technological forecasting. In view of the well-established unreliability of many such forecasts, it would seem unwarranted to be so confident in one's prediction that the requisite breakthroughs will not occur in our time as to give the hypothesis that they will a probability of less than, say, 1%.

The expected utility of a 1% chance of realizing an astronomically large good could still be astronomical. But just how good would it be for (some substantial subset of) currently living people to get access to astronomical amounts of resources? The answer is not obvious. On the one hand, one might reflect that in today's world, the marginal utility for an individual of material resources declines quite rapidly once his basic needs have been met. Bill Gates' level of well-being does not seem to dramatically exceed that of many a person of much more modest means. On the other hand, advanced technologies of the sorts that would most likely be deployed by the time we could colonize the local supercluster may well provide new ways of converting resources into well-being. In particular, material resources could be used to greatly expand our mental capacities and to indefinitely prolong our subjective lifespan. And it is by no

[10] Or whatever the population is likely to be at the time when doomsday would occur.

[11] See e.g. V. Vinge, "The Coming Technological Singularity", *Whole Earth Review*, Winter issue (1993).

[12] R. A. Freitas Jr., *Nanomedicine*, Vol. 1, Georgetown, Landes Bioscience, 1999.

[13] E.g. Moravec, Kurzweil, and Vinge op. cit.; E. Drexler, *Engines of Creation*, New York, Anchor Books, 1986.

means clear that the marginal utility of extended healthspan and increased mental powers must be sharply declining above some level. If there is no such decline in marginal utility, we have to conclude that the expected utility to *current* individuals[14] of successful colonization of our supercluster is astronomically great, and this conclusion holds even if one gives a fairly low probability to that outcome. A long shot it may be, but for an expected utility maximizer, the benefit of living for perhaps billions of subjective years with greatly expanded capacities under fantastically favorable conditions could more than make up for the remote prospects of success.

Now, if these assumptions are made, what follows about how a person-affecting utilitarian should act? Clearly, avoiding existential calamities is important, not just because it would truncate the natural lifespan of six billion or so people, but also – and given the assumptions this is an even weightier consideration – because it would extinguish the chance that current people have of reaping the enormous benefits of eventual colonization. However, by contrast to the total utilitarian, the person-affecting utilitarian would have to balance this goal with another equally important desideratum, namely that of maximizing the chances of current people surviving to benefit from the colonization. For the person-affecting utilitarian, it is not enough that humankind survives to colonize; it is crucial that extant people be saved. This should lead her to emphasize speed of technological development, since the rapid arrival of advanced technology would surely be needed to help current people stay alive until the fruits of colonization could be harvested. If the goal of speed conflicts with the goal of global safety, the total utilitarian should always opt to maximize safety, but the person-affecting utilitarian would have to balance the risk of people dying of old age with the risk of them succumbing in a species-destroying catastrophe. Mixed ethical views, which also incorporate non-utilitarian elements, might or might not yield one of these bottom lines depending on the nature of what is added.[15]

[14] On the person-affecting view, the relevant factor is not the temporal location of a person *per se*. In principle, effects on the well-being of a past or a future person could also be an appropriate target for moral concern. In practice, however, the effect on the well-being of past persons is likely to be relatively small or even zero (depending on which conception of well-being one adopts), and the effect (of any action that has a significant effect on the future) on not-yet-existing persons is very likely to be such that competing alternative actions would lead to separate sets of possible persons coming into existence, for example as a result of different sperms fertilizing different eggs, and thus there would typically be no future persons whose level of well-being is affected by our current actions. Our actions might affect which future persons there will be, but on the person-affecting view, such results provide no general moral reasons for action. One exception are fetuses and fertilize eggs, which are not yet persons, but which may become persons in the future, and whose lives our current actions may influence.

[15] I'm grateful for comments from John Broome, Milan Cirkovic, Roger Crisp, Robin Hanson, and James Lenman, and for the financial support of a British Academy Postdoctoral Award.