How many lives does the future hold?

Toby Newberry

Global Priorities Institute | May 2021

GPI Technical Report No . T2-2021

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

How many lives does the future hold?[1*]

1. Introduction

The total number of people who have ever lived, across the entire human past, has been estimated at around 100 billion.[2] The total number of people who *will* ever live, across the entire human future, is unknown - but not immune to the tools of rational inquiry. This report estimates the expected size of the future, as measured in units of 'human-life-equivalents' (henceforth: 'lives'). The task is a daunting one, and the aim here is not to be the final word on this subject. Instead, this report aspires to two more modest aims. First, to provide a robust defense of the claim that 'the future is vast in expectation': even on the conservative assumption that humanity's future remains Earth-bound over the long-term, this report estimates that it can be expected to hold at least 100 trillion lives. The report's second aim is to present an earnest, if somewhat more speculative, estimate of the true size of the future. This requires that we venture further from established research, meaning the associated estimate depends on several simplifying assumptions, and should be interpreted as less robust. Nevertheless, this report finds that the future is truly immense in expectation: on the order of 10^{30} lives, or more than a billion billion times the total number of people who have ever lived.

The remainder of Section 1 sets out the report's structure, background, and scope. Sections 2-4 form the main body of the report, and divide the central estimate into four distinct scenarios. Section 2 considers the scenario in which humanity remains Earth-bound over the long term. Section 3 considers three scenarios involving space settlement, distinguished by whether the human future is bounded by the size of the Solar System (3.1), the Milky Way (3.2), or only by known physical limits (3.3). For each of the scenarios in Sections 2-3, the relevant estimate is further subdivided into an estimate of duration, and an estimate of the average number of lives per century.Section 4 then introduces the prospect of digital persons, which can be glossed as 'the possibility that some fraction of humanity's morally-relevant successors will be instantiated digitally, rather than biologically'. Together, Sections 1-4 lay the groundwork for a simple Monte Carlo model, which outputs an estimate of the expected number of future lives, and in which readers are invited to submit their own degrees of credence in each scenario. Section 5 addresses Brandon Carter's Doomsday Argument, which sits outside of the central estimate, but which potentially impacts its outcome. Section 6 concludes.

Perhaps surprisingly, there are no serious pre-existing attempts to estimate the expected number of future lives. While there is a wealth of existing research that is directly relevant to the estimate, and which provides many individual pieces of the overall puzzle, no such piece constitutes a proper precedent.[3] To take one example, the United Nations (2019) has made projections of expected global population, but these extend only until the end of this century. Perhaps the closest thing to a precedent occurs in Nick Bostrom's *Superintelligence*, where Bostrom estimates that at least 10^{35} human lives could be created over the entire future, given known physical limits. He further estimates that at least 10^{58} human lives could be created if we allow for the possibility of digital persons (Bostrom, 2014, 101-2).[4] However, these figures

[1*] I am grateful to Will Macaskill, Toby Ord, Carl Shulman, Matthew van der Merwe, and Anders Sandberg for comments on an earlier version of this report. Nevertheless, this is still a working draft, and likely contains errors (for which I take full responsibility). Please send helpful comments to tobias.newberry@philosophy.ox.ac.uk.

[2] This estimate comes from Kaneda & Haub (2020).

[3] This includes work in ecology on the carrying capacity of the Earth and the life-span of the biosphere, work in philosophy on existential risk, and work in physics and engineering on the viability and limits of space settlement. Much of this features in the discussion to follow.

[4] This estimate is an update to earlier work (Bostrom, 2003).

fall short of representing a proper precedent in two ways. First, they concern how many people *could* exist, rather than how many people will exist in expectation. Second, they are intended to be extremely conservative. Among other things, Bostrom assumes that each star system can support an average of one billion lives per century - a figure that is smaller than almost all estimates of Earth's carrying capacity, and many orders of magnitude less than what our Solar System could plausibly support. As a result, Bostrom's figures are primarily useful in Section 4, as an estimate of the relative resource-efficiency of digital, rather than biological, lives.

The scope of this report is limited in two important ways. First, it is limited in terms of the number of scenarios considered. In principle, one could divide up possible futures in an arbitrarily fine-grained manner, theoretically yielding continued improvements to the quality of the estimate.[5] However, the aim here is just to capture those scenarios with the greatest impact on expected value. As a rule, then, any scenario not discussed below is considered out of scope, unless there is a compelling reason to think it would significantly alter the overall estimate.[6] Second, the report is limited in which questions about post-humanity it aims to address. In particular, the report explicitly assumes that 'human-life-equivalents' will remain a meaningful unit, morally speaking, over cosmological timescales.[7] As a result, most questions about which specific kinds of post-humanity appear plausible, and the grounds for thinking this unit will apply in each case, are considered out of scope.

2. Earth

Suppose humanity remains Earth-bound over the long-term: how many lives can the future be expected to hold? To answer this question, we first estimate the duration of our species' expected lifespan on Earth (in centuries), before estimating the average expected number of lives per century. The duration estimate is further divided into two sets of considerations: those concerning our 'natural' lifespan, and those concerning our 'artificial', or 'all-things-considered' lifespan. This section first summarises the evidence around humanity's natural lifespan, then assesses how artificial considerations affect this, before turning to the estimate of lives per century, and combining these to estimate the overall size of the future in this scenario.

Using geological evidence to determine the current age of our species, Snyder-Beattie et al (2019) estimate an extreme upper bound of 1 in 14,000[8] as the background rate of human extinction risk per year, and suggest that the true rate is probably less than 1 in 87,000. They demonstrate that if the rate of risk were any higher than this, it would be extremely surprising to discover that humans have been around for as long as we have.[9] These figures predict that humanity can expect a further 87,000 years on Earth without a natural extinction event occurring, and imply that recorded history represents at most 6% of humanity's total expected natural lifespan. The authors then corroborate their estimates with four further sources of

[5] For example, one could consider the scenario in which space settlement is bounded by the size of the local supercluster. This would theoretically improve the overall estimate quality, on the reasonable assumption that more narrowly specified scenarios can be estimated with greater accuracy.
[6] For example, 'what about scenarios involving digital persons?' would have been a reasonable objection had the report omitted Section 4. On the other hand, 'what about scenarios where space settlement is bounded by the size of Mars (say, because Earth is destroyed and further settlement proves unfeasible)?' appears out of scope.
[7] By contrast, 'human lives' will plausibly lose relevance over long enough timescales, perhaps as a result of evolution by natural selection, or else through other transformative processes.
[8] More precisely, they estimate only a 1 in a million chance that the annual risk is greater than 1 in 14,000, relative to a rate of 1 in 10^8.
[9] One might worry that this reasoning is vulnerable to observer selection bias. However, Snyder-Beattie et al consider these sorts of effects, and further demonstrate that their estimates are unlikely to be severely biased in this way.

empirical evidence: the average lifespans of mammalian species (1-10 million years) and hominins (~970 thousand years)[10], as well as the frequencies of mass extinction events (1 in every 30-100 million years), and natural catastrophes that might plausibly trigger human extinction[11] (from 1 in 17,000 years to less than 1 in a billion years years[12]). This supplementary evidence is consistent with 87,000 years as a lower bound on humanity's expected lifespan on Earth, and, considered as a whole, suggests that the true expected value may exceed this by several orders of magnitude. A reasonable estimate of this true natural lifespan, accounting for all of the factors given above, is 1 million years (Ord, 2020, 379).

Of course, this estimate is limited in at least two important ways. On the one hand, it ignores the risk of humanity authoring its own destruction - something that could dramatically shorten our species' lifespan. On the other hand, it ignores the prospect that we might develop tools to safeguard our future, and find solutions to most or all of the risks we face on Earth. In light of these limitations, we might think that humanity's all-things-considered lifespan is either longer or shorter than the estimate above.

For the first limitation, experts have put the total risk of existential catastrophe (including those of our own making) at 30% over the next 500 years (Leslie, 1996, 146-7), at least 25% over the long-term (Bostrom, 2002), around 17% over the next century (Ord, 2020, 167), and around 50% over the long term (Ord, 2020, 169). These individual predictions are broadly in keeping with existing aggregations. For example, a survey of participants at the Global Catastrophic Risk Conference in Oxford, 2008, found a median estimate of 19% as the overall risk of humanextinction prior to 2100 (Sandberg & Bostrom, 2008). Somewhat more conservatively, the prediction platform Metaculus lists the median probability of 'human extinction by 2100' at 1%, based on more than 700 individual estimates (Metaculus, 2021). Overall, then, there is broad agreement that the risk of human extinction in the next 100 years most likely falls somewhere between 1 and 20%, with experts favouring estimates in the upper end of that range. This is high enough to make extinction risk a central problem of our time, but, as we shall see, not so high as to have an outsized impact on humanity's expected lifespan.

For the second limitation, there are clear precedents for humanity having reduced risk in this way, such as via existing asteroid detection measures.[13] It is also true that the higher level aim of 'achieving existential security', or getting to a place where existential risk is very low, and stays very low, appears viable (Ord, 2020, 198-91). However, the crucial point to note is that these two limitations have a strikingly asymmetric effect on expected lifespan. At most, a high estimate of anthropogenic risk could reduce humanity's expected lifespan from 1 million years (based on the background rate of risk) to 0 years. By contrast, considerations pushing in the other direction could extend our lifespan on Earth by hundreds of millions of years, perhaps until the end of the biosphere (inapproximately one billion years), or even beyond. Because of this asymmetry, one would need implausibly high confidence in the chance of near-term existential catastrophe, relative to long-term existential security, for these 'artificial considerations' to imply an abbreviation of humanity's expected lifespan. On balance, then, we can treat the natural lifespan

---

[10] Note that this narrower reference class is not necessarily more reliable than the mammalian species, since it is confounded by the central role our own species played in the extinctions of many other hominins.

[11] Mass extinctions are defined by the fossil record. 'Natural catastrophes' include supervolcanoes and large meteor impacts, among other events. The frequency of such catastrophes can be estimated independently of their impacts on living things.

[12] The figure of 1 in 17,000 years refers to the risk of (smaller) supervolcanic eruptions. There may seem to be a tension between this number and the overall estimate of extinction risk: how can an extinction-level event happen once in every 87,000 years, when there is a supervolcanic eruption once in every 14,000 years (on average)? However, the authors note that only larger supervolcanic eruptions plausibly threaten human extinction, in which case none of these risk frequencies comes close to the upper bound.

[13] See, for example, Newberry (2021).

as a conservative estimate of the all-things-considered lifespan, while noting that the human future on Earth could well be much shorter, but also has a significant chance of being substantially longer.

Alongside an estimate of duration, the scale of humanity'sfuture depends on how many lives per century Earth can sustain over the long term. The UN Department of Economic and Social Affairs projects that global population will reach approximately 11 billion people by the year 2100. This provides a conservative estimate of the number of lives per centuryover the long run, on the assumption that world population remains steady at this level indefinitely. At first blush, this assumption might appear implausible: historically, the world population has not been well-characterised as remaining steady at any level.[14] On balance, however, the 11 billion figure appears conservative. While published estimates of Earth's environmental carrying capacity range as low as two billion, almost all are over 4 billion, and some range into the hundreds or thousands of billions (reflecting the possibility that technology could facilitate significantly higher population levels on Earth). It is therefore safe to treat 11 billion as a conservative lower bound on the average number of lives per century over the remainder of humanity's expected lifespan, while noting the possibility for significantly higher figures.

Putting together the estimates of at least 11 billion lives per century with the expected natural lifespan of 1 million years gives us a prediction that, in expectation, the future holds at least 100 trillion lives. This is already a thousand times more lives than everyone who has ever lived, and, as we shall see, hardly approaches the upper limits of what the human future is likely to hold.

3. Space settlement

We now consider three scenarios where humanity eventually expands beyond Earth. In these scenarios, the number of future lives is limited not by what this planet can support on its own, but by the resources of the Solar System, Milky Way galaxy, or Affectable Universe. For each scenario, the analysis consists of three parts. First, a brief argument that one should have at least non-negligible credence in the relevant scenario.[15] Second, an estimate of the scenario's implications for humanity's expected lifespan. Third, an estimate of its implications for the average number of lives per century.

3.1 The Solar System

The balance of scientific evidence suggests that settling the Solar System is already technologically viable, at least in some minimal form. To date, we have successfully sent unmanned craft to a wide range of locations throughout the Solar System, and proposals to begin Martian settlement over the next several decades have been seriously considered by both government agencies and private companies.[16] However, the scenario under consideration here requires that we go further than this minimal form. Specifically, it requires that humanity settles the Solar System up to the limit of its carrying capacity (this is the same assumption made in the Earth-bound case). While it is reasonable to be somewhat less confident in this eventuality, our existing achievements nonetheless provide evidence in its favour, and to the best of our current knowledge there are no hard barriers imposed by physics that preclude Solar System settlement of this sort. Further, the claim that it will *eventually* be viable is a modest one: denying it would require strong evidence that this will remain out of reach over the coming centuries and millennia. It would therefore be overconfident to hold negligible credence in this scenario.

[14] With the possible exception of some times in prehistory.
[15] In one sense, these arguments are not necessary, since readers are free to input their own credences in the model.
[16] SpaceX is actively pursuing a proposal of this sort (see, e.g. Drake, 2016).

Unfortunately, many of the methods used to estimate humanity's expected life-span on Earth break down when we consider space settlement. In particular, reference class forecasts based on the lifespans of other mammalian or hominin species are significantly less relevant, because none of these other species expanded beyond Earth. Estimates based on the incidence of mass extinction events, and the frequency of natural catastrophes, are also less relevant: all known categories of mass extinction, and almost all natural catastrophes, have effects that are confined to Earth. While none of this is to say that reasonable estimates of humanity's expected life-span cannot be made for these scenarios, it is true that the evidential base gets somewhat thinner as we consider grander scenarios. To reflect this, the estimates below typically consist of an upper and lower bound, together with a simplifying assumption about the distribution of possibilities, as opposed to an overall 'best guess'.

In the case of the Solar System, there are at least three considerations that might inform the estimate. First, since Solar System settlement does not preclude humanity continuing on Earth, we can take the earlier figure of one million years as a reasonable lower bound on expected lifespan in this scenario. Second, the main lifetime sequence of the Sun will last for another five billion years, and can serve as a conservative upper bound on expected lifespan.[17] Third, there are several known natural risks that could have Solar System-wide effects: gamma-ray bursts, nearby supernovae, and the risk that the vacuum of space itself might collapse. However, each of these risks is relatively well-understood, and all are sufficiently small as to have virtually no impact on the overall estimate. For example, the risk of vacuum collapse has been well-characterised at no higher than 1 in 10 million per century (Tegmark and Bostrom, 2005), with best-guess estimates suggesting it is as low as 1 in 10^{600} per year (Buttazzo et al., 2013). It therefore represents a life expectancy of at least one billion years, and has a negligible overall effect on the estimate here. Together, these considerations imply that humanity's expected lifespan in this scenario will fall between one million and five billion years. If we assume the relevant probability density is uniformly distributed across this interval, we arrive at an average life expectancy of around 2.5 billion years.

The carrying capacity of the Solar System is most plausibly limited by available energy, which provides an estimate of the expected number of lives per century in this scenario.[18] The total power output of the Sun is approximately two billion times that which sustains all life on Earth, and represents a highly conservative estimate of the total power available to a technologically mature civilisation.[19] Combining this with the earlier estimate for Earth's carrying capacity, which was itself highly conservative, we produce an estimate of around 20 quintillion biological humans for the carrying capacity of the Solar System.

When multiplied by the earlier estimate of humanity's expected lifespan, conditional on settling the Solar System (2.5 billion years, or 25 million centuries), this figure (20 quintillion lives per century) produces an estimate of around 10^{27} lives for the expected size of the future. Even for someone who considers substantial Solar System settlement to be a one in a million shot, this amounts to a thousand billion billion expected future lives.[20]

---

[17] In the same way that the natural end of the biosphere does not necessarily spell the end of life on Earth, it is possible that a civilisation capable of surviving until the end of the Sun's main lifetime sequence would be capable of surviving into the red giant phase. In this sense, the five billion year figure is conservative.

[18] It is tempting to think that the carrying capacity would also be limited by the current availability of terrestrial habitats; for instance, the surface areas of Mars, Venus, and the rocky moons. In fact, there are strong reasons to favour the construction of off-planet habitats: construction is more energy-efficient outside of gravity-wells, and in the absence of atmospheric weather.

[19] It is conservative in that it uses a figure for the total Solar energy incident on Earth, only a small fraction of which goes towards sustaining human life. Moreover, only a small fraction of the Sun's total mass-energy is released as light.

[20] This calculation assumes that the rate of settlement is 'fast' relative to humanity's expected lifespan, such that the average number of lives per century, in the long run, approximately equals the carrying capacity. However, given that

3.2 The Milky Way

Conditional on Solar System settlement being achieved, it appears highly plausible that humanity could expand to other star systems across the Milky Way. To motivate this claim, consider the fact that by the 1970s we had already reached an important milestone in interstellar transit: sending the Voyager I and II probes beyond our Solar System, and into interstellar space. The Voyager probes also used considerably weaker propulsion mechanisms than current technology permits[21], and do not reflect the cutting edge of what humanity could achieve as regards interstellar transit. It would therefore seem overconfident to believe that humanity will never expand beyond our Solar System, even given the fullness of time. Furthermore, because this scenario assumes Solar System settlement, the viability of interstellar transit is essentially all it needs to get off the ground: a civilisation that is capable of settling one Solar System would plausibly be able to settle other systems, provided it can get to them.

Once more, the increase in the scale of the scenario comes with a corresponding increase in both expected duration and expected carrying capacity. With respect to duration, we can make a similar series of moves to those in the preceding sub-section. First, we can take the expected duration in the case of Solar System settlement as a lower bound on the expected duration here. The question of a reasonable upper bound is more difficult, since the Milky Way does not have as well-defined a lifetime sequence as the Sun. Here, we use the end of the stelliferous era - the time after which very few new stars form - to play this role. This will happen in approximately ten trillion years (Adams & Laughlin, 1999). However, it is worth noting that the end of the stelliferous era is not really a good proxy for the maximum possible lifespan of a galaxy-wide civilisation.[22] Nevertheless, its inclusion here permits us to make illustrative estimates, while reflecting the facts that (1) simple and effective proxies for this upper bound are few and far between, and (2) the upper bound can be expected to fall somewhere between the upper bounds for scenarios 3.1 and 3.3. Similar to Solar System settlement, there are very few known natural risks that could cause extinction at an interstellar scale. However, because of the greater timescales involved, the risk of vacuum collapse is potentially relevant to the estimate here. Based on the evidence cited in the previous section, a reasonable assumption is that there is at least a 50% chance that vacuum collapse either does not exist, or else operates on timescales even longer than those considered here. This consideration can therefore be treated as approximately halving the duration estimate. If we again assume a uniform distribution, these numbers result in an average expected duration of around 2.5 trillion years.

For carrying capacity, the calculation is quite a bit simpler: we can simply take the estimate for this Solar System and multiply it by the number of stars in the Milky Way.[23] This gives an estimate of around 10^{25} lives per century as the carrying capacity of our galaxy. Together, these figures produce an estimate of around 10^{35} lives for the expected size of the future, conditional on humanity settling the Milky Way.

---

we are talking about billions of years, 'fast' could in this context mean 'over the course of hundred of thousands of years'.

[21] For example, it has long been known that nuclear power could produce enough thrust to reach speeds measured in percentage points of the speed of light - several orders of magnitude faster than the voyager probes (see, e.g. Dyson, 1968).

[22] One obvious weakness here is that a galaxy-wide civilisation could plausibly survive beyond the end of the stelliferous era: there would still be stars even when star formation ceases, and it may be possible to derive additional energy in other ways (see, for example, footnote 30, below).

[23] Between 100 billion and 400 billion. Two other points may also be worth noting. First, the Milky Way will collide with the Andromeda galaxy in some 4.5 billion years, adding a further ~trillion stars to those within reach of an interstellar civilisation. Second, this method is limited by the fact that not all stars are identical to the Sun.

3.3 The Affectable Universe

Moving from interstellar to intergalactic settlement involves another step-change in scale. While we have no precedent for sending material over intergalactic distances, recent work demonstrates that this is achievable using standard physics (Sandberg and Armstrong, 2013).[24] Theoretical evidence of this sort, even from experts, should weigh less than the more practical evidence for interstellar transit, but it would still seem unduly sceptical to believe that intergalactic travel is close to impossible.

However, it does not seem likely that settlement beyond the Milky Way would meaningfully impact the duration of humanity's expected lifespan. As in the previous scenario, the end of the stelliferous era marks a plausible upper bound to this lifespan - and the end of the stelliferous era is not confined to the Milky Way.

Of course, the prospect of settling other galaxiesdoes have significant implications in terms of carrying capacity. Our Local Group, to which the Milky Way belongs, contains around 80 galaxies or dwarf galaxies, most of which hold many millions of stars.At the most extreme limits, the 'affectable universe' - the portion of the universe within our power to impact causally[25] - contains around 20 billion galaxies (Ord, 2020, p233). As a result, we might expect an intergalactic civilisation to support hundreds, millions, or even billions times more lives per century than one confined to a single galaxy. To arrive at an estimate of carrying capacity, we use the size of our LocalGroup as a lower bound on the number of galaxies humanity might reach, with the total affectable universe as an upper bound, and again assume a uniform distribution.[26] This gives an estimate of around 4 billion galaxies, conditional on humanity achieving intergalactic settlement. If we then assume that each of these galaxies could support, on average, as many lives as the Milky Way (10^{25} per century), we arrive at an estimated carrying capacity of around 10^{35} lives per century for the affectable universe.[27]

Multiplying this figure (10^{35}) by the estimated same expected duration as in the Milky Way scenario (50 trillion years) gives an expected 10^{46} lives for the total size of the future.

4. Digital persons

So far, we have estimated the expected number of future lives in four different scenarios. Importantly, these are estimates of the expected number of *biological* lives: among other things, they are based on evidence about the number of present-day humans the Earth can support. In this section, we turn to consider the prospect of digital persons. Following the same pattern as above, we begin by motivating the scenario's plausibility, before addressing its implications for the expected size of the future. However, digital persons merit a slightly different treatment in terms of how the prospect fits into the overall model. For the scenarios addressed in Sections 2 and 3, each successive scenario (approximately) entailed all previous scenarios, such that one's credence over the full set should sum to 1. By contrast, we treat digital

---

[24] That is, without exploiting speculative assumptions, e.g Einstein-Rosen wormholes.
[25] This region is sometimes referred to as the universe's 'Event Horizon'.
[26] This is a simplification. One way it could be improved upon is by assuming a uniform distribution over 'civilisational expansion speed', rather than over the number of galaxies reached (this would reduce the expected number of galaxies by about a half). One could also make further refinements to the speed distribution, drawing on available information about which physical barriers might place hard limits on this. However, improvements in this direction seem unlikely to alter the estimate by more than an order of magnitude, and so have been omitted here (see, for a partial explanation, the first limitation mentioned in the introduction).
[27] As with stars, there is considerable variation in the size and composition of galaxies. This means that the simple multiplicative method used here is only informative as a broad illustration of scale, and should not be read as authoritative.

persons as an independent prospect, which would act as a constant multiplier to the estimates in each other scenario. In other words, the assumption is that the development of digital persons is more or less independent of different levels of space settlement.[28]

There is a consensus among experts about the viability of humanity eventually creating artificial general intelligence, and even broad agreement that the relevant timescale is years or decades, rather than centuries or millennia. A survey of more than 300 machine learning experts, carried out in 2016, asked about the timeframe until an AI system would be 'able to accomplish every task better and more cheaply than human workers'. On average, the experts predicted a 10 percent chance of this happening as soon as 2025, and a 50 percent chance of it happening by 2061. These timeframes are corroborated by a separate survey from 2014, where machine learning experts estimated 2075 as the median year by which this level of AI development would most likely be achieved.[29] The second survey also recorded the proportion of experts who think humanity will *never* achieve this level of AI development, and found this to be a small minority, at only 16.5%.[30] In light of this, it would be unreasonable to hold very low levels of credence in this prospect, especially over the long term.

Of course, the creation of artificial general intelligence does not necessarily amount to digital persons, in that the relevant digital entities need not be 'people' (morally speaking). We should therefore treat the prospect of digital persons as somewhat less likely than artificial general intelligence. However, in the absence of compelling reasons for thinking digital persons are impossible, the fact that experts agree on the viability of artificial general intelligence, with relatively high confidence, seems sufficient to motivate non-negligible credence in this scenario.

If a significant fraction of humanity's morally-relevant successors were instantiated digitally, rather than biologically, this would have truly staggering implications for the expected size of the future. As noted earlier, Bostrom (2014) estimates that 10^{35} human lives could be created over the entire future, given known physical limits, and that 10^{58} human lives could be created if we allow for the possibility of digital persons. While these figures were not intended to indicate a simple scaling law[31], they do imply that digital persons can in principle be far, far more resource efficient than biological life. Bostrom's estimate of the number of digital lives is also conservative, in that it assumes all such lives will be emulations of human minds; it is by no means clear that whole-brain emulation represents the upper limit of what could be achieved. For a simple example, one can readily imagine digital persons that are similar to whole-brain emulations, but engineered so as to minimise waste energy, thereby increasing resource efficiency.

Here, we include estimates of the 'digital carrying capacity' for each of the scenarios considered so far - in other words, the number of digital persons we might expect the Earth, Solar System, Milky Way, or Affectable Universe to support. These should not be understood as authoritative, but rather are intended to further illustrate the extraordinary impact that digital persons could have on the expected number of future lives. In the model linked below, we invite readers to input their own estimates of the relative resource efficiency of digital versus biological life, as well as the fraction of humanity's future resources they expect to support digital entities. For each of the scenarios in Sections 2-3, we then multiply the resource fraction by the expected number of future lives, and scale the outcome of this by the efficiency

[28] If anything, the development of digital persons would simplify space settlement (in the same way that unmanned space missions are simpler than those with crews). In this sense, the assumption of independence can be regarded as highly conservative.
[29] Technically, they were estimating the year in which there is a 90% chance of the existence of machine intelligence "that can carry out most human professions at least as well as a typical human."
[30] Again, this is technically the fraction of respondents who thought 'there is no year by which I am 90% sure humanity will have achieved this level of AI development.'
[31] I.e. that $10^{58}/10^{35} = 10^{23}$ digital lives can be supported for each biological life, at any scale.

estimate. Finally, the resulting product is added back to the expected number of biological lives (minus those that have been 'converted' to digital entities). This gives an estimate of the total number of lives, both digital and biological, for each scenario.

How many digital lives might the Earth support? One way of estimating this is to assume that solar energy could be converted to digital consciousness at approximately the same rate that energy from food is converted to biological consciousness in the human brain (i.e. about 20 watts). On Earth, the total average solar irradiance at surface level is on the order of 1,730 terawatts. Making the conservative assumption that 1% of this might be available to support digital persons, we arrive at an estimated carrying capacity of around 140 trillion lives.[32]

Applying this same method to the other scenarios, we get similar increases in scale. For the Solar System, we can take 1% of the Sun's *total* power output (~10^{24} watts) as our starting figure. Dividing this by 20 gives a digital carrying capacity of around 10^{23} lives. For the Milky Way, we can simply scale this by the number of stars in our galaxy, which gives a digital carrying capacity of around 10^{34} lives. Finally, we can scale this by the number of galaxies in the Affectable Universe (8 billion), which gives a digital carrying capacity of close to 10^{44} lives. Especially for the latter two scenarios, these figures are conservative in that they ignore the prospect of energy from sources other than starlight. As Ord (MS) notes, only a tiny fraction (less than 0.1%) of stellar mass will eventually be radiated as starlight, and there may be other ways that technologically advanced civilizations could generate energy.[33]

5. The Doomsday Argument

One might worry that all of the above reasoning can be undone on purely probabilistic grounds, by appeal to something like the Doomsday Argument. There are at least three points worth making about such a move, which, taken together, suggest that it does not succeed in undercutting the case presented above. In essence, the Doomsday Argument makes the point that it would be highly unusual to find oneself in a highly unlikely position: if all people who will ever be born were ordered by their date of birth, 90% of them would find themselves in the first 90% of people. By the same token, we should think there is only a small chance we are in the first small percentage of people who will ever be born - implying that the chance of there being many, many more people born in future is equivalently small.

The first point to note about this is that it remains highly contentious: anthropic reasoning is an unsettled field in general, and there is no consensus that the Doomsday Argument, in particular, succeeds. In fact, the argument remains beset by proposed rebuttals, and it is not at all clear it will escape this continued onslaught unscathed. To take one example, the 'Self Indication Assumption'[34] states that, *ceteris paribus*, an observer should reason as if they are randomly selected from the set of all *possible* observers. As an attack on the Doomsday Argument, we can gloss this as the claim that 'if more people are born overall, there is a greater chance I will be among them'. Assuming Bayesianism, the size of this effect precisely cancels out the effect of the Doomsday Argument, suggesting the argument provides no information whatsoever about the total number of people who will ever exist.

Second, notwithstanding the argument's name, the actual prediction it makes is that the expected number of people who will ever live is infinite, given infinite time. Mathematically, the argument's central claim

[32] I.e. This is 1% of 170,000 terrawats, divided by 12.
[33] For example, feeding matter into black holes may be a far more efficient way of accessing energy.
[34] This principle was first proposed by Dennis Dieks (1992), and given its current name by Nick Bostrom (2002). It has been defended by Bartha & Hitchcock (1999), and, more recently, Ken Olum (2001).

implies that the eventual population, measured in multiples of the number of people who have lived so far, will fit a Pareto distribution: such a distribution has an infinite mean (Ord, MS).

Third, the Doomsday Argument remains, at most, a way of setting a prior: it claims that, in the absence of other evidence, one can estimate the total number of people who will ever live based on one's likely birth position. But, as the reasoning above should make clear, it is far from the only evidence available. Overall, these considerations imply that most aspects of the Doomsday Argument, including its basic philosophical soundness, and the direction it points in this particular case, remain in question. It should not be treated as a strong objection to the foregoing arguments.

6. Conclusion

The preceding considerations provide the foundations for a simple Monte Carlo model. Table 1 summarises the various estimates of duration and carrying capacity that are used in the model, and defended above, using my best guess estimates of expected duration and expected carrying capacity in each case. Table 2 presents three further model parameters that do not fit cleanly into the same framework, including the relative efficiency of digital versus biological life. Finally, Table 3 summarises the outputs of the model for a person who is sceptical of the arguments above. In particular, this person thinks that humanity stands no chance whatsoever of settling the affectable universe or the Milky Way, and only a 1 in 10,000 chance of settling the Solar System. Further, they estimate that only a trillionth of humanity's future resources will go towards supporting digital persons. Even with these sceptical estimates in place, the expected future is almost unimaginably vast: extending over millions or billions of years, and containing some 10^{28}- that is, 10 billion billion billion - lives.

Scenario	**Duration (centuries)**	**Carrying capacity (lives per century)**	**How many lives does the future hold?**
Earth (Homo sapiens reference class)	10^{3}	10^{10}	10^{13}
Earth (hominin reference class)	10^{4}	10^{10}	10^{14}
Earth (mammalian reference class)	10^{4}	10^{10}	10^{14}
Earth (mass extinctions reference class)	10^{6}	10^{10}	10^{16}
Earth (digital persons)	10^{4}	10^{14}	10^{18}
The Solar System	10^{8}	10^{19}	10^{27}
The Solar System (digital persons)	10^{7}	10^{23}	10^{30}
The Milky Way	10^{11}	10^{25}	10^{36}
The Milky Way	10^{11}	10^{34}	10^{45}

(digital persons)			
The Affectable Universe	10^{10}	10^{35}	10^{45}
The Affectable Universe (digital persons)	10^{10}	10^{44}	10^{54}

Table 1: core estimates[35]

Parameter	**Estimate**
Number of stars in the Milky Way	10^{11}
Number of galaxies humanity could potentially reach	10^{3} - 10^{10}

Table 2: misc. parameters

Inputs for different scenarios	**'An extremely conservative reader'**
Credence in scenario 2: Earth	~1
Credence in scenario 3.1: The Solar System	10^{-4}
Credence in scenario 3.2: The Milky Way	0
Credence in scenario 3.3: The Affectable Universe	0
Fraction of humanity's future resources devoted to scenario 4: digital persons	10^{-12}
Resource efficiency of digital lives relative to biological lives	10^{9}
How many lives does the future hold?	**10^{28}**

Table 3: Example estimate

Ultimately, it is the fact that humanity's *potential* is so remarkably vast that drives the estimate of expected value. In terms of duration, it is at least *prima facie* possible that humanity could last until the end of Earth's biosphere, the death of the Sun (~5 billion years), or the death of the last stars (~10 trillion years). In terms of population, it is possible that the number of people alive at any one time could far exceed present levels, especially when taking into account the prospects of digital persons, and of humanity continuing beyond Earth. Crucially, the argument that our future is vast in expectation requires only a small degree of credence in such expansive possibilities. By contrast, the claim that our future will be

[35] Estimates in excess of 1 quadrillion are rounded to the nearest order of magnitude.

anything shorter than a million years, or contain any fewer than 100 trillion lives, in expectation, would require both extraordinary scepticism about these grander futures, and extraordinarily high confidence that humanity will cause its own destruction in relatively short order.

7. References

Bostrom, N. (2002). 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards.' *Journal of Evolution and Technology, 9.*

Bostrom, N. (2003). 'Astronomical Waste: The Opportunity Cost of Delayed Technological Development.' *Utilitas*, Vol. 15, No. 3.

Bostrom, N. (2014). *Superintelligence.* Oxford: Oxford University Press.

Buttazzo, D. et al. (2013). 'Investigating the Near-Criticality of the Higgs Boson.' *Journal of High Energy Physics*, 2013(12).

Kaneda, T., Haub, C. (2020). 'How many people have ever lived on Earth?', *Population Reference Bureau.*

Leslie, J. (1996). *The End of the World: The Scienceand Ethics of Human Extinction.*Routledge.

Metaculus (2017). 'Will humans go extinct by 2100?', accessed 4/5/2021 https://www.metaculus.com/questions/578/human-extinction-by-2100/.

Müller, V. & Bostrom, N. (2016). 'Future Progress in Artificial Intelligence: A Survey of Expert Opinion' in *Fundamental Issues of Artificial Intelligence*, Müller, V. ed., Synthese Library, Berlin: Springer.

Ord, T. (2020). *The Precipice.* London: Bloomsbury.

Ord, T. (MS). 'The edges of our universe'. arXiv:2104.01191

Sandberg, A. & Bostrom, N. (2008): "Global Catastrophic Risks Survey", Technical Report #2008-1, Future of Humanity Institute, Oxford University.

Snyder-Beattie, A., Ord, T., and Bonsall, M. (2019). 'An upper bound for the background rate of human extinction,' *Sci Rep* 9, 11054.

Tegmark, M. & Bostrom, N. (2005). 'Is a Doomsday Catastrophe Likely?' *Nature*, 438(7069).

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Volume I: Comprehensive Tables.