Short-Term vs. Long-Term EA Focus Depends on Annual Extinction Probability

Magnus Vinding

23-11-17

An Integral Favoring Short-Term Focus?

Discussions of the overwhelming importance of the far future of our civilization often miss factoring in the probability of such a future existing in the first place. When we assume certain plausible estimates of that probability, we find that the importance of the next 100-10,000 years dominate in expectation.

Let $P(E)$ be the probability that civilization has of going extinct per year. Let $p(t)$ be the expected number of morally relevant individuals in our civilization at any given time $t =$ "number of years from now". The expected population at any given time can thus be defined as

$$EP_t = (1 - P(E))^t p(t)$$

while the expected number of morally relevant individual life-years that will exist from now until some time T is given by:

$$ELY_T = \int_0^T (1 - P(E))^t p(t) dt$$

Playing with specific values of $P(E)$ and $p(t)$ yields some interesting results. Plausible values seem to suggest that we should focus on the short term. At least if we are trying to maximize quality of individual life-years rather than to maximize the probability of them existing.

For instance, we may start by assuming a constant population, $p(t) = K$. The exact value is not important. We then get

$$ELY_{T,K} = \int_0^T (1 - P(E))^t K dt$$

Let us now assume that the probability civilization has of going extinct each year is a mere one percent. We thus get the following specific values of $ELY_{T,K}$

$$ELY_{10,K} = K \cdot 9.51$$

$$ELY_{50,K} = K \cdot 39.30$$

$$ELY_{100,K} = K \cdot 63.08$$

$$ELY_{200,K} = K \cdot 86.17$$

$$ELY_{300,K} = K \cdot 94.62$$

$$ELY_{400,K} = K \cdot 97.71$$

$$ELY_{500,K} = K \cdot 98.85$$

$$ELY_{1,000,K} = K \cdot 99.49$$

$$ELY_{10,000,K} = K \cdot 99.50$$

$$ELY_{T\to\infty,K} = K \cdot 99.50$$

In other words, on these two assumptions, $P(E) = 0.01$ and $p(t) = K$, the expected life-years that will exist in our civilization from now and until a hundred years from now is roughly two thirds of the expected life-years (in our civilization) from now and to infinity, while the expected life-years from today to 300 hundred years from now is almost the same as from now and to infinity.

What if the probability of extinction each year were only 0.1 percent? Then we get

$$ELY_{10,K} = K \cdot 9.95$$

$$ELY_{100,K} = K \cdot 95.16$$

$$ELY_{1,000,K} = K \cdot 631.99$$

$$ELY_{10,000,K} = K \cdot 999.45$$

$$ELY_{T\to\infty,K} = K \cdot 999.5$$

As we can see, the picture changes, but not in a fundamental way. The expected number of life-years from now and until a thousand years from now is about two thirds of what we should expect from now and until infinity on this assumption, and the latter is roughly equal to $ELY_{10,000,K}$. The bulk of expected life-years still lies surprisingly close to today.

"But", one may object, "the assumption about constant population is clearly implausible, as the number of morally relevant individuals in our civilization is most likely going to increase in the future."
Whether this is truly "most likely"is, I think, far from clear, yet even if we assume that the population will indeed grow rapidly, this actually does not change the basic conclusion drawn above.

Assume we colonize the galaxy with the speed of light. The volume of a growing sphere grows with its radius to the third power. Thus, this is presumably the maximum growth rate of the population of a colonizing civilization (at least it will be eventually):

$$p(t)_{max-growth} = Ct^3$$

where C is some growth constant (one could argue we should add a constant representing the starting population, as we otherwise get $p(0) = 0$, yet this is not relevant here, as the absence of this constant only serves to strengthen the steelman we are currently building for focusing on the far future, and because this starting population quickly becomes negligible in this scenario).

This gives us the following expected individual life-years integral

$$ELY_{p(t)_{mg},T} = \int_0^T (1 - P(E))^t C t^3 dt$$

Assuming $P(E) = 0.001$, we now get

$$ELY_{10,C} = 2480 \cdot C$$

$$ELY_{50,C} = 1.5 \cdot 10^6 \cdot C$$

$$ELY_{100,C} = 2.3 \cdot 10^7 \cdot C$$

$$ELY_{1,000,C} = 1.1 \cdot 10^{11} \cdot C$$

$$ELY_{10,000,C} = 5.9 \cdot 10^{12} \cdot C$$

$$ELY_{100,000,C} = 6.0 \cdot 10^{12} \cdot C$$

$$ELY_{T \to \infty,C} = 6.0 \cdot 10^{12} \cdot C$$

So as in the case with constant population and $P(E) = 0.001$, we again, even given maximum population growth, have virtually no expected life-years after 10,000 years, although the difference between $ELY_{10,000}$ and $ELY_{1,000}$, as well as the ELY at all other times earlier than a thousand years, is considerably greater in this case compared to the case $p(t) = K$.

If we modeled more realistically so that this cubic growth started happening, say, a hundred years from now rather than now, this would make it conditional on our getting there, which reduces its probability and the ELY significantly. Especially if one modeled it with $P(E) = 0.01$ in the next hundred years, which would be roughly consistent with estimates of the probability of extinction occurring the coming century.

But will the probability of extinction then not drop as we colonize? Perhaps. Yet it could also increase, as larger, less centralized systems also become more vulnerable, at least in some ways, cf. Phil Torres' "Why We Should Think Twice About Colonizing Space". So it is not clear.

Given the vastly greater number of expected life-years in the scenarios where there is a low extinction probability and large population growth, these scenarios should arguably still dominate our considerations and be what we seek to influence the most. Yet the consideration presented here does serve to significantly dampen the extent to which long-term influence dominates short-term influence.

In sum, the far future may, depending on our estimates, be unlikely to ever occur, yet its expected size implies that impacting it should plausibly still be considered our main priority. The consideration outlined here suggests, however, that the effects our actions have on the long-term future may not dominate their short-term effects as much in expectation as a naive analysis that ignores extinction probability would conclude. Indeed, if one then also factors in the consideration that the long-term future is difficult to influence reliably, it does not seem implausible that we should focus roughly equally on impacting the short-term and long-term future to have the best impact in expectation.