The Risk that Humans Will Soon Be Extinct

JOHN LESLIE

Abstract
If it survives for a little longer, the human race will probably start to spread across its galaxy. Germ warfare, though, or environmental collapse or many another factor might shortly drive humans to extinction. Are they likely to avoid it? Well, suppose they spread across the galaxy. Of all humans who would ever have been born, maybe only one in a hundred thousand would have lived as early as you. If, in contrast, humans soon became extinct then because of the population explosion you would have been 'fairly ordinary'. Roughly ten per cent of all humans would have been your contemporaries. Now (as the cosmologist Brandon Carter saw to his dismay) a scientific principle tells us not to treat observations as highly extraordinary when they could easily be fairly ordinary. How to apply the principle is controversial, yet it seems we can safely conclude that humanity's chances of galactic colonization cannot be high. Still, we should work to make them as high as possible, resisting those philosophers who argue that human extinction would be no tragedy.

Scenarios in which the human race ends fairly soon are easy to construct. Many are quite plausible. The shortest plausible scenario inserts under a second between the start and the end of the doomsday process. [Refuse to believe it? Read onwards!] However, the process could perhaps take years, decades or centuries. Consider a major exchange of nuclear bombs, for instance. There still exist over twenty thousand of them. Thanks to Russia's financial troubles the threat that nuclear Armageddon will begin by accident is probably worse now than during the Cold War. Nitrogen oxides from a nuclear holocaust could ruin the ozone layer, Earth's shield against ultraviolet light. The radioactive fallout would work mischief. A nuclear winter, severe cooling lasting for months, might result from all the soot which burning cities and forests threw into the atmosphere. The human race might be exterminated, perhaps (Carl Sagan suggested) through deaths of micro-organisms crucial to the health of the biosphere. And since Nevil Shute's novel of 1957, *On the Beach*, people have grown used to the idea that the extermination could be fairly gradual.

doi:10.1017/S0031819110000446
© The Royal Institute of Philosophy, 2010

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

Look next at the possibility of utterly disastrous climate change, a greenhouse-effect runaway. The geological record reveals that major jumps in temperature sometimes happen quite quickly. Now, to get the consensus needed for persuading the politicians in Rio in 1992 the International Panel on Climate Change disregarded worst-case predictions and latest-available evidence. It even dealt with biological feedback loops in just one sentence: 'Biological feedbacks have not yet been taken into account.' Has the Panel since changed its ways? There is little sign of it. Politicians demand findings that are very uncontroversial and the IPCC, remember, was created to provide exactly such findings. But scenarios involving runaway overheating are readily available, biological feedback loops often playing crucial roles. For instance: (i) Ocean waters warm up, becoming less able to absorb man-made carbon dioxide, the factor chiefly responsible for the change; (ii) waters rich in nutrients rise to the warmed sea surface less often so that phytoplankton grow more slowly, absorb less carbon dioxide and generate less dimethyl sulphide, a substance which encourages the birth of the clouds that cool us in daytime; (iii) many phytoplankton die because carbon dioxide has acidified the oceans; (iv) hotter weather increases production of carbon dioxide by plants and soil microbes; (v) tundra melt and peat bogs dry out, producing yet more carbon dioxide and vast amounts of another greenhouse gas, methane, molecule for molecule perhaps thirty times as powerful; (vi) resultant changes in high altitude clouds make them trap more heat; (vii) drought then kills vegetation, returning carbon dioxide to the atmosphere; (viii) next, the ravages of methane and other greenhouse gases deplete the hydroxyls which are so important in destroying those gases; (ix) there follows a retreat of sea ice so that less sunlight is reflected back into space; (x) heating of the oceans thereupon releases trillions of tons of methane which are at present locked up in the clathrates of the continental shelves; (xi) the new heat produces much more water vapour, an extremely important greenhouse gas, so that a greenhouse runaway occurs. For advanced life forms, Earth becomes uninhabitable.

Why did nothing so awful happen during the Cretaceous era, the last time the atmosphere was thick with greenhouse gases? Perhaps only because the sun was cooler then. Over Earth's history solar luminosity has been rising gradually, by a total of about thirty per cent. Earth during the Cretaceous could already have been close to a runaway. Maybe we could get one shortly, the grand die-off beginning in the tropics and developing over several decades.

Alternatively, disaster might strike first in North America. It might start with a computer virus ruining the Internet that now controls so

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

much of what keeps developed nations from starving. Well armed brigands from the stricken regions could descend like locusts on all food supplies elsewhere.

The spatiotemporal details of any extinction process would be of great ethical significance. The tens of millions of deaths in the Second World War and how they were distributed among the warring parties would be far less significant, I'd say, than whether half a billion humans managed to live onwards for a century when an unstoppable slide towards extinction had begun, and whether the process started with the swift extinction of those nations which had done least to cause the fatal nuclear exchange, greenhouse calamity or whatever. All the same, I suggest the following. Whether the extinction process took only a fraction of a second, spreading outward at almost the speed of light from a single centre (once again read onwards), or was instead a matter of many decades and much geographical complexity, would be far less important than the fact that humans would be dying out before beginning serious efforts to spread across their galaxy.

This is because such efforts would probably succeed; because the number of human lives which would then be lived would probably be in the hundreds of trillions, assuming that our much modified descendants could be counted as 'humans'; and because most of the lives would probably be more worth living than ones in our technologically primitive era. The weightiest of all ethical imperatives could be Spread Through The Galaxy! The extinction of the human race before galactic colonization could begin would be immensely tragic because of the vastly many worthwhile lives that would then have no possibility of being lived.

Many a professional philosopher has suggested, in contrast, that the loss of a future containing countless richly satisfactory human lives would be hardly any loss at all. Its drawback would be simply that folk would have hoped in vain that the future would contain such lives. Its ugliness, that's to say, would lie in its non-fulfilment of things wished by the dead, as when computer error sends Aunt Jane's fortune to the orphanage, not the Home for Abandoned Cats she recommended to her executors. Just what, the philosophers inquire, could anyone have lost through not being born, for *just who* wouldn't have been born? People possess identities only after coming into the world, they protest, and only people with identities can lose anything.

Note, though, that the very same philosophers declare (a) that it would be wicked for people to give birth if they could be confident that those born (whoever they were, so let's here forget the little

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

matter of their identities!) would have lives that were very miserable, and (b) that generating such lives would be more evil, the greater the number of those lives. But, I ask, isn't there too weird an asymmetry here? Little or no duty, say our philosophers, to generate humans likely to have lives richly worth living even if you could generate trillions of them by snapping your fingers, because they wouldn't have identities until generated, *and yet* there's a firm obligation *not* to generate humans likely to be miserable, and above all not to generate many of them. Well, to accept that sort of thing you must perhaps be a professional philosopher.

It seems to me, anyhow, that the duty to attempt galactic colonization is immensely strong-assuming, that's to say, that there'd not be trillions of extraterrestrials in danger of extermination or enslavement. Might humankind plausibly become extinct before the process could get started? Galactic colonization will probably begin within a couple of centuries if we don't annihilate ourselves during that period. What chance, then, does the human race stand of surviving for another two hundred years?

Comets and asteroids are highly unlikely to annihilate everybody during the next two hundred years. During the next two hundred years Earth could well be struck by a body large enough to wipe out several cities if it hit an ocean, producing a gigantic tsunami. Yet even something as massive as the object that killed off the dinosaurs would find it hard to exterminate all humans, and things so huge arrive only about once in a hundred million years. Similarly we need have little fear of volcanic super-eruptions. They occur only rarely. And while a wandering black hole may conceivably have flattened the forests of Tunguska, events of this exotic sort couldn't be frequent enough to concern us.

Germ warfare, in contrast, could very soon be extraordinarily dangerous. Scientists could produce new diseases that killed far more efficiently than the Spanish 'flu of 1918 which ended more lives than a World War had just done. An aggressor's attempts at self-protection might be ineffective, in which case maybe every single human would be killed. Would nobody be evil enough to risk it? The world contains some deeply unpleasant persons and, now that mammalian cells can be grown on tiny beads, a single bottle can produce germs in numbers which previously required large factories. Maybe some filthy little dictator is growing germs secretly, or some madman (from a world containing some eight hundred billionaires) with enough cash for hiring experts, or some terrorist (or even some 'biohacker') willing to risk eliminating everyone. Here is one of numerous fields where modern science and

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

technology are putting ever greater destructive power into the hands of ever more individuals. Unless we can all be trained or monitored very successfully, 'Brainwashing' and 'Police State' to many liberals, we must expect frightful calamities sooner or later.

It is often population pressures that lead to war and terrorism, and the world's population is still exploding. There are roughly seven billion humans today, meaning very little farmland for each. There could be twelve billion by the end of the century. Hence even without war and terrorism the environment could come under catastrophic pressure. Perhaps it already has, owing to the unholy alliance between pesticides and fertilizers; the cutting down of forests; the loss of biodiversity; the large 'dead zones' where fish used to breed, coastal sectors whose waters are low in oxygen because of agricultural runoffs and sewage outflows; the chemicals whose thinning of the ozone layer has still by no means ended. Earth is suffering heavily through pollution quite apart from any resulting climate change.

Next, consider naturally occurring diseases. Cramming of half the world into giant cities, continued growth of international travel, and weakening of immune systems through onslaughts from the ninety thousand chemicals which are synthesized on an industrial scale, could combine to produce plagues of startling killing power. Genetic analysis suggests that the AIDS virus originated as early as 1884, only starting to spread swiftly when the modern world arrived. Maybe luck alone explains why its present triumphs are not larger. The closely related visna virus, which infects sheep, is spread simply by coughing. Hesitate before protesting that humankind has survived all diseases for many millennia, for you might then be forgetting Observational Selection. There are some very lethal diseases. Untreated rabies, for instance, is always fatal. Perhaps the universe is richly supplied with planets on which utterly lethal diseases have destroyed all intelligent observers; but had Earth been one such planet, would your observations show that you lived there?

Before continuing down the list of possible disasters let me make clear that my book in the area, though it carries the potentially alarming title *The End of the World: the science and ethics of human extinction*,[1] is actually none too gloomy. While the human race seems pretty

[1] Leslie, *The End of the World* (London and New York: Routledge, 1996). For interesting recent papers on the risk of human extinction, see *Global Catastrophic Risks*, (eds.) N. Bostrom and M. M. Circovik (Oxford: Oxford University Press, 2008). See also Bill Joy, 'Why the Future Doesn't Need Us', *Wired* (April, 2000), 238–262: 'Our most powerful 21st-century

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

sure to end sooner or later instead of continuing onwards for infinite time, the book gives it a good chance, perhaps (says page 146) as high as seventy percent, of surviving the next five centuries, after which it could have fairly good prospects of spreading right across its galaxy. The book therefore contrasts markedly with the one by Martin Rees, the Astronomer Royal, published in Britain as *Our Final Century*[2] and in North America as *Our Final Hour*. Rees' 'hour' is of course metaphorical. His 'century', however, is literal. As estimated by Rees we have only about a half chance of getting through the next ninety years. But while considerably less pessimistic than he is, even I can see that humanity's future is disturbingly uncertain. Bear in mind Enrico Fermi's question of why we have failed to find extraterrestrials. Studies suggest that once a species had experienced its Industrial Revolution it could spread across its galaxy in a few million years. Well, our galaxy has existed some thousand times as long, and it contains many billion stars much like the sun. Even if only a small proportion have suitable planets, why hasn't a single planet given birth to intelligent beings who have spread sufficiently for us to detect them? Maybe intelligent species appear quite often but then destroy themselves by their technological breakthroughs.

One of our latest technologies, genetic engineering, has alarmed many people. Most experts call its risks fairly small but social pressures might be influencing them. Instead of just affecting industry as in the case of regulations applying to nuclear power plants, efforts to restrict genetic engineering threaten the salaries and research grants of numerous scientists. I lose little sleep worrying about the 'green scum disaster' in which a genetically engineered super-organism wipes out everything else, yet maybe not even this can be ruled out. And it was definitely a shock to find Toronto's *Globe and Mail* reporting on its front page (but with no mention of any risk) that researchers had used genetically modified salmonella bacteria to produce harmless, temporary infections able to act as contraceptives that rendered recipient women infertile for months. Aren't salmonella bacteria extremely widespread, and mightn't the genetically

---

technologies – robotics, genetic engineering, and nanotech – are threatening to make humans an endangered species' is the summary at the head of this influential article. Global warming is discussed expertly by James Hansen, director of the NASA Goddard Institute for Space Studies, in his *Storms of My Grandchildren: The Truth About the Coming Climate Catastrophe and Our Last Chance to Save Humanity* (New York: Bloomsbury, 2009).

[2] M. Rees, *Our Final Century* (London: Heinemann, 2003).

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

modified ones mutate so that they spread efficiently from woman to woman, infecting many women repeatedly? What mightn't *that* do to humanity's survival prospects? Consider also that genetic engineers could develop terrible germ warfare agents. The greatest threat to the human race could be from these, perhaps let loose by terrorists. Australian genetic engineers managed to modify a mousepox so that it killed every infected mouse. They did it by mistake. Think what might be done by actual planning.

A variant on any green scum disaster is the 'gray goo calamity' featuring devices first proposed by Richard Feynman, 'nanomachines' no larger than bacteria. Controlled by internal computers, these could be self-reproducing. After all, bacteria can be viewed as natural machines and they reproduce themselves with ease. If it used sunlight as its power source and naturally available chemicals, self-reproducing nanomachinery might perhaps 'reduce the biosphere to dust in a matter of days' (Eric Drexler's words). While this might be unlikely to occur by sheer accident, because free-ranging nanobots could be far harder to design and less commercially useful than ones confined to factories, the disaster could be initiated by hobbyists, terrorists, apocalyptic sects, criminals, madmen, haters of humanity, or disciples of Schopenhauer (see below). Probably nothing of the sort could happen during the next two hundred years yet it is difficult to be sure.

Risks connected with computers of ordinary dimensions could materialize rather sooner. Whereas a collapse of the Internet might be survived easily enough, a solar storm like that of 1859 could knock out power grids for weeks or even months, putting any computer-dependent regions at risk. It is hard to feel convinced that folk in those regions would all sit at home and starve instead of plundering the non-computer-dependent regions. Again, people have speculated that something grim could happen as nations gave more and more of their decision-making to computers of ever greater intelligence. The most powerful nations, winners in the struggle to survive, could have computers programmed simply to defeat other nations economically and militarily. Their countries might have few citizens, almost everything being done by machine. A conceivable finale would be that computers replaced humans entirely.

Some variants on this scenario have the computers replacing us through sheer cunning. We create them without first solving the tricky problem of how to make them reliably friendly towards us. They then figure out how to increase their own intelligence very quickly, after which they run circles round our safety measures. Other variants, though, have them taking over with our full blessing.

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

A leading expert on artificial intelligence, Hans Moravec, said in his book *Mind Children*[3] that he believed the human race was in its final hundred years and that he intended to work for its speedy extinction. Instructed to design improved versions of themselves, computers would become immensely more intelligent than humans. They would be virtually immortal. Their thought processes would be far less neurotic. In a word, they would be finer, happier beings than us. We ought to be eager to have them replace us.

Moravec here assumed that computers could become not only very intelligent but also fully conscious. Many a philosopher would agree, yet many another wouldn't. My own view is that devices much like the ordinary digital computers of today could soon be much cleverer than humans but would nevertheless lack consciousness of any worthwhile kind. Worthwhile consciousness comes only with what Roger Penrose has termed *globality*, the remarkable unity that characterizes our conscious states. Here elements join together in immensely complex patterns but each individual element is rather obviously 'just an abstraction' in a crucial sense. It is no more capable of existing *all by itself* than the blue and the red elements of a pebble's purpleness, which are of course real but not in the solid way that the pebble is real. Our states of mind (or, since psychologists have long known that the human mind isn't as fully unified as Descartes believed, at least various sub-states of our states of mind) possess Penrose's 'globality' not because we have immaterial souls but because areas in the human brain carry out quantum computations: computations of the 'highly parallel' kind, in principle able to consider hugely many possibilities simultaneously, which are performed by quantum computers. Mohan Sarovar and Gregory Scholes found in 2009 that complex, fairly long-lived quantum phenomena (complex quantum superpositions interfering constructively with themselves, and states entangled over distances large by quantum standards) are used by green bacteria and by marine algae to help capture and process sunlight: so much for the common argument, rightly opposed by Penrose, that such phenomena would of course be disrupted by the 'excessive heat' of biological systems! Quantum computers (could we but construct them sufficiently skilfully before immensely intelligent, virtually immortal, non-neurotic digital computers replaced us) might indeed have worthwhile consciousness because richly complex quantum wholes are not structures whose every component has an existence clearly separate from that of

[3] H. Moravec, *Mind Children* (Cambridge, Mass.: Harvard University Press, 1998).

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

its neighbours.[4] A quantum computer's consciousness could be real like that of a human, instead of like the so-called 'collective consciousness' of an angry mob or the consciousness that some ascribe to ant colonies or immune systems because these perform complex information-processing. However, all this is controversial.

Another controversial issue is whether physicists could destroy the world by experiments at very high energies. It has been discussed in the physics journals. Reassuring arguments have been crafted by safety committees. While joining me in thinking that the experiments of the next few decades are likely to cause nothing frightful, you could still be worried by the degree of courage needed by any committee which contemplated saying that some particle accelerator costing many billion dollars might perhaps be world-destroying. The physics involved could be so difficult and a risk of disaster estimated at anything below one or two percent could seem so trivial, particularly to people uninterested in all the humans who wouldn't be born if things went wrong. The potentially disturbing truth is that energies reached inside particle accelerators have increased roughly tenfold in each decade since 1900. Now, in the comparatively cool world that succeeded the Big Bang the most concentrated releases of energy, presumably unthreatening because nothing disastrous has ever happened because of them, result from head-on collisions between cosmic rays, particles which sometimes pack the punch of rifle bullets. What if physicists of future years did better than that? They well could. One of the world's leading physicists, Steven Weinberg, suggests in his *Dreams of a Final Theory*[5] that through using very powerful lasers to accelerate particles even Planck-scale energies might be had. He comments that a collision between two particles at those energies would discharge into an extremely tiny volume the power of a head-on collision not just of bullets but of small jet aircrafts. Could an approach to this sort of thing be dangerous?

Possibly so. According to most physicists what we call empty space is in fact filled with one or more force fields, 'scalar fields' vital to the world as we know it. And (maybe as early as in this century, unlikely though most experts think it) experimenters could reach energies above those of all cosmic ray collisions. They might then produce a tiny bubble of new-strength scalar field. The bubble would keep expanding for reasons like those which keep a rock tumbling downhill.

[4] I discuss these things in *The End of the World* and in chapter two of *Infinite Minds* (Oxford University Press: Clarendon Press, 2001).

[5] S. Weinberg, *Dreams of a Final Theory* (London: Hutchinson, 1993). See 187–188 in particular.

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

It would expand at nearly the speed of light. A small fraction of a second later, Earth would have been destroyed; shortly afterwards, the whole solar system; after roughly a hundred thousand years, our entire galaxy; and so on.

People often dismiss the risk of any such disaster, pointing to an article in which Rees and Piet Hut discussed cosmic ray collision energies.[6] Yet in his *Before the Beginning*[7] Rees calls the danger real enough, given how ingenious physicists are at pushing accelerator energies upwards. 'Caution,' he comments, 'should surely be urged (if not enforced) on experiments that create energy conditions that may never have occurred naturally.' His *Our Final Century* adds that if accelerators 'became a hundred times more powerful' then fears of a world-destroying Final Experiment could well be warranted 'unless in the meantime our understanding has advanced enough to allow us to make firmer and more reassuring predictions from theory alone'.[8] Already the Large Hadron Collider features beams with the kinetic energies of Eurostar trains moving at full speed, beams squeezed down to widths of sixteen microns at their collision points. Well, when (even if only as something to be expected less than once a decade although slight changes in the degree of squeezing might at times increase the particle collision rate considerably, as might any momentarily increased bunching of the particles) very unusually many particle collisions chanced to occur in very specially close spatiotemporal proximity then the outcome could be significantly different from what happens on those rare occasions when two cosmic rays of the highest energies smash into each other. Moreover creation of numerous sub-critical bubbles could involve a risk that some bubble would grow catastrophically by exploiting quantum uncertainties. Such a disaster 'may not be likely, but it only takes one event' as Marc Sher said in an early article that remains important today. It cited four hundred and sixty-five other studies but commented that all could be 'swept into the dustbin' by some new finding.[9]

Toby Ord and colleagues at Oxford's Future of Humanity Institute pointed out recently that if a risk is estimated as very small then the estimate can be as good as meaningless if there's an

[6] Rees and Hut, 'How stable is our vacuum?', *Nature* **302** (April, 1983), 508–509.

[7] Rees, *Before the Beginning* (Reading, Mass.: Addison-Wesley, 1997). The quotation is from page 207.

[8] Rees, Our Final Century (London: Heinemann, 2003), 122.

[9] M. Sher, 'Electroweak Higgs potentials and vacuum stability', *Physics Reports* **5** and **6** (1989), 273–418; see especially 335–336.

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

appreciably greater risk that the reasoning behind it is wrong. What would you think of the general who said, 'We believe that enemy gun to be fifty miles away, giving it only a one in a billion probability of hitting you', while failing to lay stress on the one in five hundred chance that the gun was half a mile away instead? Notice in this connection that arguments for saying the Large Hadron Collider could never create an Earth-eating black hole were changed repeatedly. In 1999 it was declared confidently that the Collider could create no black holes. When 2003 revealed that String Theory's extra dimensions might permit their creation, it was held with equal confidence that any that were created would 'evaporate' at once through their Hawking radiation. When more work showed that this might be mistaken, it was argued instead that white dwarf stars showed no signs of being eaten by the black holes that could confidently be expected there. Earth-eating black holes continued to strike me as highly unlikely but you can see why others became more and more uneasy.

In connection with Weinberg's idea that lasers could be used to accelerate particles, note that gigantic lasers have already delivered for extremely brief periods more power than that flowing from all the world's power stations combined. And the hard-X-ray lasers of Ronald Reagan's Strategic Defense Initiative were designed to do better even than that, in the instants before being vaporized by the small nuclear bombs that would have powered them.[10]

A possible ground for taking doomsday scenarios seriously was discovered by the cosmologist Brandon Carter. It has now been defended by several people. I have countered a large number of objections which others have thrown at it.[11] Suppose that a hundred intelligent races of more or less the same size will eventually have evolved in our universe. We couldn't at all expect to be in *the very earliest*, could we? Now, very similarly, you and I couldn't expect to be among the very earliest, for instance in the first one

[10] N. Brown, *New Strategy Through Space* (Leicester: Leicester University Press, 1990); see especially page 85.

[11] In *The End of the World*, which also includes much detailed discussion of many doomsday scenarios (it even mentions – for storm clouds could be seen approaching even in 1996 – a tiny risk associated with world-wide economic collapse, 'the huge U.S. governmental, commercial and household debt' and 'stock-market trading in derivative securities' combining to 'vaporize financial markets'); in several articles, for instance 'Time and the anthropic principle', *Mind* **101** (1992) 521–540; in 12–14 of *Immortality Defended* (Oxford: Blackwell, 2007); and during a tour as the British Academy-Royal Society of Canada Exchange Lecturer (I thank the British Academy for financing it).

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

percent, of all humans who will ever have existed. Yet this is where we'd have been, were the human race to survive for many more centuries even at just its present size. If it spread right across its galaxy then we could easily have been in the first thousandth of one percent. But if, in contrast, humans became extinct shortly then you and I would have lived at the same time as up to ten percent of all humans ever. Out of all humans so far, fairly many are alive today thanks to the population explosion. Well, quite a useful principle of scientific reasoning runs as follows. If your observations could plausibly be viewed as fairly ordinary, don't hurry to classify them as highly extraordinary! Now consider your own position inside the total spatiotemporal spread of the human race. To view it as fairly ordinary rather than highly extraordinary, you need only think that humans will quite soon be extinct.

Imagine a scene towards the end of this century. Ten billion humans walk the Earth but all will die inside a month because of germ warfare. One of the doomed complains of his atrocious luck in having been born so late. 'There have been many thousand generations since the start of human history yet here I am in the very last generation, the one which gets wiped out by the progress of military science. What fantastically unusual misfortune!' Now, mightn't that be foolish? Roughly ten percent of all humans ever entering the world would be alive at that very moment. How could it be 'fantastically unusual' to find yourself in company with ten percent of all humans?

Here is another way of looking at this kind of thing. Suppose that intelligent life is bound to evolve many times in our universe of over a hundred billion galaxies. If intelligent species found it easy to survive long enough to spread through their galaxies then wouldn't almost all intelligent beings live at times when their races *had* spread through their galaxies, because at those times the races in question would contain so immensely many inhabitants? But do you and I find ourselves in a race which has spread through its galaxy?

People often protest that we necessarily find ourselves alive in this present period rather than another. If the period is one when the human race hasn't managed to spread through its galaxy, then so what? Finding ourselves alive now cannot suggest that there won't be many trillion humans later, they maintain. Now is *now* and not later, and anyone who wasn't born when you were wouldn't be *you*. It's absurd, their protest runs, to ask what period you could expect to have been born into, against the background of this or that story about how humans are distributed in time. 'When you could expect to have been born' is a nonsensical notion! But any protest on these

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

lines is simply wrong, I suggest. Suppose you have forgotten your birthday. Is it likely to be in March? No, for only comparatively few humans have March birthdays. Or consider lemmings. Where does a typical lemming find itself? At a time when there are hardly any lemmings, or after a lemming population explosion?

What, though, if you grew dissatisfied with viewing yourself just as somebody alive at present? What if you changed to classifying yourself as having in addition been born on exactly such and such a date, for instance January the first 1950? In the scenario where the human race soon becomes extinct, your new 'narrowing of the reference class' (as probability theorists would call it) might at first strike you as making a very significant difference. Whereas you'd formerly have been in, perhaps, the one-in-ten category among all humans past, present and future, you would now be in a category several thousand times smaller. However, you might then notice that your new approach would once again treat you as more statistically unusual than before, and with precisely the same increased degree of statistical unusualness, in the scenario where the human race instead spreads across its galaxy. Perhaps this second scenario used to put you in the one-in-a-billion category. This could sound a very tiny category. But the reference-class narrowing would move you into something a lot more tiny: more tiny by precisely the same thousands of times. It would therefore have no influence on the crucial point. It wouldn't at all change how much more statistically exceptional the second scenario would make you.

Brandon Carter saw that reflections on those lines ought to increase any doubts we might have about a long future for humankind, a future perhaps containing many hundred trillion humans scattered among the stars. This has come to be known as 'the doomsday argument'. After looking at the dangers confronting the human race but before considering Carter's line of reasoning I'd have estimated that humans had only about a five percent chance of becoming extinct during the next two hundred years. How does Carter's argument affect my estimate? It makes it about six times more pessimistic. What, though, if there were hugely many extraterrestrials distributed through time and through all those billions of galaxies? Wouldn't an observer's chance of being *a human in the early twenty-first century* be virtually unaffected by whether most *human* observers lived much later? Unfortunately Carter's logic survives this question. For we ought (until we find enough evidence to the contrary) to try to see ourselves as 'fairly ordinary' inside as many as possible of the various classes into which we fall. Now, we fall not only into the class of *observers* but also into that of *human observers*. And we

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

humans of the early twenty-first century could be *fairly ordinary among human observers* were the human race destined to end soon.

I can still have strong doubts about the variant on Carter's logic that was discovered quite independently by Richard Gott.[12] Gott tends to argue simply that your chance and mine of being in, for instance, the first 2.5 percent of all humans who will ever have lived *is 2.5 percent*, and that's that. If, when human history ended, all humans had always guessed they lived in the first 2.5 percent then exactly 2.5 percent of them would have guessed correctly, he points out. Yet this is too simple an argument, surely. We have evidence of actual risks facing the human race and of attempts to counter them by, for example, trying to stop filthy dictators developing biological weaponry. Surely we shouldn't disregard such evidence. Carter, at any rate, intended his argument to be just a magnifying glass for increasing whatever pessimism people had developed after thinking about filthy dictators, a greenhouse runaway, nuclear bombs and other threats. Suppose that contemplating all the threats still left you extremely confident in a long future for humankind. Then, although reducing your confidence, the doomsday argument might fail to remove it. Pessimism which starts off extremely small can remain small even when magnified.

Another important point is this. Carter's doomsday argument can operate in a smooth mathematical way only if ours is a fully deterministic world, a world in which the number of humans who will ever have lived is therefore something which has already been fixed: something 'already out there' in the realm of genuine truths. In that case your situation and mine could be compared to that of a man watching names being drawn from a hat. He knows his name was in the hat only once, and that the hat contained either ten names or else a thousand. He has just seen his name, as the seventh to be drawn. He now asks how likely it is that the hat contained ten names only. He ought to take into account two things: first, how confident he was initially that the hat contained a thousand names, and second that the arrival of his name on the seventh draw would have been more probable had there been ten names instead. A mathematical formula, Bayes' Rule, can tell him what he should now think. But this formula works smoothly only because the number of names in the hat was all along something 'already there'.

You might reasonably believe that truths about the future of humankind are not already there in any comparable way. The number

[12] R. Gott, 'Implications of the Copernican principle for our future prospects', *Nature* **363** (May, 1993), 315–319.

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

of humans who will ever have lived may depend, for example, on whether some politician is going to push a nuclear warfare button. Whether the politician will push the button could in turn depend, many a theologian would argue, on some utterly free decision which the politician has yet to make. Or, which many a quantum physicist would judge important, what might be crucial is various radically indeterministic events which have yet to occur in the politician's brain. The moral is that the doomsday argument may run fairly smoothly only as a means of reducing great confidence in a long future for humans: confidence that the long future is 'there' or 'as good as there' in the sense that it is inevitable or very nearly inevitable. If you are far from having this sort of confidence then the argument perhaps oughtn't to influence you all that greatly. For example, suppose you thought the only two possibilities were these: that the human race would become extinct during this or the next century, and that it would instead come to contain many hundred trillion members through spreading right across its galaxy. Despite all the nuclear bombs etcetera, you could reasonably believe there was a fair probability, perhaps approaching thirty percent, that it would spread right across the galaxy. That's how I myself think, at least when I haven't just been reading the works of those quantum physicists who argue that even the quantum realm is ruled by rigid determinism. But had I instead felt sure the world was fully deterministic, making the number of humans who would ever have been born into something that had been settled ever since the Big Bang, then Brandon Carter's reasoning would have shaken my optimism severely. It would have made me put the chances of spreading right across the galaxy at well below one in a thousand.

What if we became convinced that the world was indeterministic in ways crucial to the future of humankind? [Think again of the politician with his finger on the nuclear button. Think of his brain as indeterministic at the quantum level.] In that case, just how optimistic could we be that future humans would exist in huge numbers? Well, we'd be exhibiting truly weird contempt for nuclear bombs, germ warfare, a greenhouse runaway and so forth, if we put humanity's chances of spreading across the galaxy at anything over fifty percent. Even in a radically indeterministic world you and I should surely shrink from judging ourselves *more likely than not* to be *very, very unusually early* among all humans who will ever have arrived in the galaxy as measured by a clock ticking whenever a new human is born, a clock whose ticks now have things like nuclear bombs in the background.

People worrying about human extinction have also raised other problems with which philosophers struggle. The chief one is this.

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

If humankind's imminent extinction seemed very possible, would it be a fact that we ought to try to prevent it?

For a start, is it ever a fact in any straightforward sense that we ought to do this or that? So-called facts about what we ought to do: mightn't these be just matters of taste? Or perhaps of what actions we prescribe, so that 'It's a fact you ought to care about future humans' says no more than 'Caring about future humans truly is what I'm hereby telling you, myself and everyone else to do'? Next there is the point touched on earlier. Could it ever be a fact that people had a right to be born? If various possible folk never got to be born because the human race had been wiped out by germ warfare, then how could they be rightly indignant about their fate? People who remain perpetually confined to the realm of the merely possible instead of being real people certainly cannot be indignant. And how can we be indignant on their behalf when they have no actual identities?

Here let me just comment that when we name various threats to the human race, *threats from faulty philosophy* should be on our list. I know of no forceful reason for denying that facts about good and bad are 'out there in the real world' just as much as the fact that two and two make four; and yet, philosophers line up to deny it. Again, you will find many philosophers who run the argument that merely possible people have no actual identities and that therefore there can be no strong duty to keep the human race in existence. Ethics, they often declare, is just a matter of reducing conflicts between people if there happen to be any people. And you will find other philosophers who consider that getting rid of humankind would be really rather a good thing because of all the unhappy humans who would live if the human race continued onwards. On this way of thinking even one miserable human per million might be too much.

You will actually find some philosophers, Schopenhauer for example, writing that almost every human is more miserable than happy so that it would be wonderful if humans went extinct.

There are difficult philosophical points in this area. Still, please don't be persuaded by the general run of the professionals who discuss them. If you were, you could find yourself declaring that a rapid end to the human race would be no disaster. Yet advanced intelligence may arrive in a galaxy only through a great deal of luck. It could even be that humans are the only markedly intelligent species in our entire universe. If they survived the next century or two then their descendants could have a fair chance of spreading right through the Milky Way and perhaps through other galaxies

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press

also. The lives of those descendants could be much more worth living than ours, and they could well exist in huge numbers. Are we really to disregard all this?

Everything might hang on whether a few billion dollars went to building artificial refuges so that a few score humans could survive germ warfare. However, almost nobody is interested in building them. The first major artificial habitat, 'Biosphere 2', cost only about a hundred and fifty million dollars. [Procurement cost plus program cost of the B-2 Spirit 'Stealth Bomber'? Per aircraft, roughly twenty times as much.] Designed with the help of sloppy science, the habitat quickly failed to be self-sustaining. Later a 'Biosphere 3' in Siberia kept three people alive for half a year. Otherwise nothing has been tried.

University of Victoria
johnlesl@uoguelph.ca

https://doi.org/10.1017/S0031819110000446 Published online by Cambridge University Press