X7 – EXchanging Worldviews, 7: EXamining Possibilities for the Human System

Dear: As promised at the end of the previous chapter, with this chapter I want to examine possibilities for humanity's future. In that regard, your grandmother recently relayed to me a relevant "one liner" that she heard from a paleontologist: "Everything we study is either extinct or evolving." Similarly for our species: either humanity will become extinct or evolve.

Of course there's the possibility that we'll do what fundamentalist religious people want, i.e., maintain the *status quo* (viz., "the state in which [we are]"), but it's highly doubtful. Yet, fundamentalists (especially Muslims) desperately cling to the delusion that they've found "perfection". Would that they'd give some thought to Stewart's idea (quoted at the end of the previous chapter) that each human has been given the marvelous opportunity to participate in evolution! I'd say to all religious people: Don't blow it!

Meanwhile, essentially all of this chapter will be "just" quotations from what two other authors recently wrote about the more likely of the two possibilities for humanity:

1) That humanity will become extinct (for which the quoted author, Nick Bostrom, "guesstimates" a probability of about 25%, ☹), and

2) That humanity will continue to evolve (and the quoted author, Clayton Naff, is so "upbeat" about this possibility that he suggests that humanity will evolve into what we would consider to be gods! ☺).

And in advance, Dear, I should acknowledge that these two quotations are long, but I hope you'll consider them carefully.

Below, the first and sobering quotation deals with the possibility of humanity's extinction. It was written by Nick Bostrom of the Department of Philosophy at Yale University (and also Oxford University).[1]

---

[1] In the 2001 version of this article (the one copied), Bostrom's affiliation was given as the Department of Philosophy at Yale University; in the 2002 version, his affiliation is given as Department of Philosophy, Oxford University. The article (available on the internet at several sites) was published in the *Journal of Evolution and Technology, Vol. 9,* March 2002.

I've quoted a substantial portion of Bostrom's article, to increase the probability that you'll read it! To make it simpler for you to read it, I've cut some material (leaving ellipses, "…") and removed his references (which account for most of the ellipses). In a few places, I've added some notes in brackets. The footnotes within the text were written by the author, Bostrom.

Existential Risks
Analyzing Human Extinction Scenarios and Related Hazards

Nick Bostrom, PhD
Faculty of Philosophy, Oxford University…

Existential risks

In this paper we shall discuss… global, terminal risks. I shall call these *existential risks*. Existential risks are distinct from global endurable risks. Examples of the latter kind include: threats to the biodiversity of Earth's ecosphere, moderate global warming, global economic recessions (even major ones), and possibly stifling cultural or religious eras such as the "dark ages", even if they were to comprise the whole global community, provided they are transitory (though see the section on "Shrieks" below). To say that a particular global risk is endurable is evidently not to say that it is acceptable or not very serious. A world war fought with conventional weapons or a Nazi-style *Reich* lasting for a decade would be extremely horrible events even though they would fall under the rubric of endurable global risks since humanity could eventually recover. (On the other hand, they could be a terminal *local* risk for many individuals and for persecuted ethnic groups.)

I shall use the following definition of existential risks:

> *Existential risk* – One where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.

An existential risk is one where humankind as a whole [is] imperiled. Existential disasters have major adverse consequences for the course of human civilization for all time to come.

The unique challenge of existential risks

Risks in this… category are a recent phenomenon. This is part of the reason why it is useful to distinguish them from other risks. We have not evolved mechanisms, biologically or culturally, for managing such risks. Our intuitions and coping strategies have been shaped by our long experience with risks such as those of dangerous animals, hostile individuals or tribes, poisonous foods, automobile accidents, Chernobyl, Bhopal, volcano eruptions, earthquakes, droughts, World War I, World War II, epidemics of influenza, smallpox, black plague, and AIDS. These types of disasters have occurred many times and our cultural attitudes towards risk

have been shaped by trial-and-error [and "trial and success"!] in managing such hazards. But tragic as such events are to the people immediately affected, in the big picture of things – from the perspective of humankind as a whole – even the worst of these catastrophes are mere ripples on the surface of the great sea of life. They haven't significantly affected the total amount of human suffering or happiness or determined the long-term fate or our species.

With the exception of a species-destroying comet or asteroid impact, which is very infrequent event, there were probably no significant existential risks in human history until the mid-twentieth century, and certainly none that it was within our power to do something about. The first manmade existential risk was the detonation of the first atomic bomb. At the time, there was some slight concern that the explosion might start some kind of runaway chain-reaction by "igniting" the atmosphere. Although we now know that such an outcome was physically impossible, it qualifies as an existential risk that was present at that time. In order for there to be a risk it suffices that there is some *subjective probability*, given the knowledge and understanding available, of an adverse outcome, even if it later turns out that objectively there was no chance of something bad happening. *If we don't know whether something is objectively risky or not, then it is risky in the subjective sense*. The subjective sense is of course what we must base our decisions on.[2] At any given time we must use *our best current subjective estimate* of what the objective risk factors are…

A much greater existential risk emerged with the build-up of nuclear arsenals in the US and the USSR. An all-out nuclear war was a possibility with both a substantial probability and with consequences that *might* have been persistent enough to qualify as global and terminal. There was a real worry among those best acquainted with the information available at the time that a nuclear Armageddon would occur and that it might annihilate our species or permanently destroy human civilization.[3] Russia and the US retain large nuclear arsenals that could be used in a future confrontation, either accidentally or deliberately. There is also a risk that other states may one day build up large nuclear arsenals. Note however that a smaller nuclear exchange, for instance between India and Pakistan, is not an existential risk, since it would not destroy or thwart humankind's potential permanently. Such a war might however be a local terminal risk for the cities most likely to be targeted. Unfortunately, we shall see that nuclear Armageddon and comet or asteroid strikes are mere preludes to the existential risks that we will encounter in the 21st century.

---

[2] This can be seen as the core wisdom of the so-called Precautionary Principle… Any stronger interpretation of the principle, for instance in terms of where the burden of proof lies in disputes about introducing a risky new procedure, can easily become unreasonably simplistic.

[3] President Kennedy is said to have at one point estimated the probability of a nuclear war between the US and the USSR to be "somewhere between one out of three and even "… John von Neumann (1903-1957), the eminent mathematician and one of the founders of game theory and computer science and who as chairman of the Air Force Strategic Missiles Evaluation Committee was a key architect of early US nuclear strategy is reported to have said it was "absolutely certain… that there would be a nuclear war; and… that everyone would die in it…"

The special nature of the challenged posed by existential risks is illustrated by the following points:

- Our approach to existential risks cannot be one of trial-and-error. There is no opportunity to learn from errors. The reactive approach – see what happens, limit damages, and learn from experience – is unworkable. Rather, we must take a proactive approach. This requires *foresight* to anticipate new types of threats and a willingness to take decisive *preventive action* and to bear the costs (moral and economic) of such actions.

- We cannot necessarily rely on the institutions, moral norms, social attitudes or national security policies that developed from our experience with managing other sorts of risks. Existential risks are a different kind of beast. We might find it hard to take them as seriously as we should simply because we have never yet witnessed such disasters.[4] Our collective fear-response is likely ill calibrated to the magnitude of threat.

- Reductions in existential risks are *global public goods*… and may therefore be undersupplied by the market… Existential risks are a menace for everybody and may require acting on the international plane. Respect for national sovereignty is not a legitimate excuse for failing to take countermeasures against a major existential risk.

- If we take into account the welfare of future generations, the harm done by existential risks is multiplied by another factor, the size of which depends on whether and how much we discount future benefits…

In view of its undeniable importance, it is surprising how little systematic work has been done in this area. Part of the explanation may be that many of the gravest risks stem (as we shall see) from anticipated future technologies that we have only recently begun to understand. Another part of the explanation may be the unavoidably interdisciplinary and speculative nature of the subject. And in part the neglect may also be attributable to an aversion against thinking seriously about a depressing topic. But the point is not to welter in gloom and doom but simply to take a sober look at what could go wrong so we can create responsible strategies for improving our chances of survival. In order to do that, we need to know where to focus our efforts.

---

[4] As it applies to the human species, that is. Extinction of other species is commonplace. It is estimated that 99% of all species that ever lived on Earth are extinct. We can also gain some imaginative acquaintance with existential disasters through works of fiction. Although there seems to be a bias towards happy endings, there are exceptions such as the film *Dr. Strangelove*… and Nevil Shute's poignant novel *On the Beach*… Moreover, in the case of some existential risks (e.g. species-destroying meteor impact) we do have experience of milder versions thereof (e.g., impacts by smaller meteors) that helps us quantify the probability of the larger event. But for most of the serious existential risks, there is no precedent.

Classification of existential risks

We shall use the following four categories to classify existential risks:[5]

> *Bangs* – Earth-originating intelligent life goes extinct in relatively sudden disaster resulting from either an accident or a deliberate act of destruction.
>
> *Crunches* – Scenarios where the potential of humankind to develop into posthumanity[6] is permanently thwarted although human life in some form continues.
>
> *Shrieks* – Some form of posthumanity is attained but it is an extremely narrow band of what is possible and desirable.
>
> *Whimpers* – A posthuman civilization arises but evolves in a direction that leads gradually but irrevocably to either the complete disappearance of the things we value or to a state where these things are realized to a minuscule degree of [what] could have been achieved.

Armed with this taxonomy, we can begin to analyze the most likely scenarios in each category. The definitions will also be clarified as we proceed.

Bangs

This is the most obvious kind of existential risk. It is conceptually very easy to understand. Here are some of possible ways for the world to end in a bang… I have tried to rank them roughly in order of how probable they are, in my current personal estimation, to cause the extinction of Earth-originating intelligent life; but my intention with the ordering is more to provide a basis for further discussion than to make any firm assertions.

Deliberate misuse of nanotechnology

In a more mature form, molecular nanotechnology will enable the construction of bacterium-scale self-replicating mechanical robots that could feed on dirt and other organic matter… Such replicators can eat up the biosphere or destroy it by other means such as by poisoning it, burning it, or blocking out sunlight. A person of malicious intent in possession of this technology may be able to cause the extinction of intelligent life on Earth by releasing such nanobots into the environment.[7]

---

[5] The terminology is inspired by the famous lines of T. S. Eliot (from "The Hollow Men"): "This is the way the world ends // Not with a bang but a whimper" and also by the title of philosopher John Earman's book on the general theory of relativity…

[6] The words "posthumanity" and "posthuman civilization" are used to denote a society of technologically highly enhanced beings (with much greater intellectual and physical capacities, much longer life-spans, etc.) that we might one day be able to become…

[7] Nanotechnology, of course, also holds huge potential for benefiting medicine, the environment, and the economy in general; but that is not the side of the coin that we are studying here.

The technology to produce a destructive nanobot seems considerably easier to develop than the technology to create an effective defense against such an attack (a global nanotech immune system, an "active shield"…). It is therefore likely that there will be a period of vulnerability during which the technology must be prevented from coming into the wrong hands. Yet this technology could prove hard to regulate, since it doesn't require rare radioactive isotopes or large, easily identifiable manufacturing plants, as does production of nuclear weapons…

Even if effective defenses against a limited nanotech attack are developed before dangerous replicators are designed and acquired by suicidal regimes or terrorists, there would still be the danger of an arms race between states possessing nanotechnology. It has been argued… that molecular manufacturing would lead to both arms race instability and crisis instability, to a higher degree than was the case with nuclear weapons. Arms race instability means that there would be dominating incentives for each competitor to escalate its armaments, leading to a runaway arms race. Crisis instability means that there would be dominant incentives for striking first. Two roughly balanced rivals acquiring nanotechnology would, on this view, begin a massive buildup of armaments and weapons development programs that would continue until a crisis occurs and war breaks out, potentially causing global terminal destruction. That the arms race could have been predicted is no guarantee that an international security system will be created ahead of time to prevent this disaster from happening. The nuclear arms race between the US and the USSR was predicted but occurred nevertheless.

Nuclear holocaust

The US and Russia still have huge stockpiles of nuclear weapons. But would an all-out nuclear war really exterminate humankind? Note that: (i) For there to be an existential risk it suffices that we can't be sure that it wouldn't. (ii) The climatic effects of a large nuclear war are not well known (there is the possibility of a nuclear winter). (iii) Future arms races between other nations cannot be ruled out and these could lead to even greater arsenals than those present at the height of the cold war. The world's supply of plutonium has been increasing steadily to about two thousand tons, some ten times as much as remains tied up in warheads… (iv) Even if some humans survive the short-term effects of a nuclear war, it could lead to the collapse of civilization. A human race living under stone-age conditions may or may not be more resilient to extinction than other animal species.

We're living in a simulation and it gets shut down

A case can be made that the hypothesis that we are living in a computer simulation should be given a substantial probability… If we are, we suffer the risk that the simulation may be shut down at any time. A decision to terminate our simulation may be prompted by our actions or by exogenous factors.

Badly programmed superintelligence
When we create the first superintelligent entity… we might make a mistake and give it goals that lead it to annihilate humankind. For example, we could mistakenly elevate a subgoal to the status of a supergoal. We tell it to solve a mathematical problem, and it complies by turning all the matter in the solar system into a giant calculating devise, in the process killing the person who asked the question…[8]

Genetically engineered biological agent
With the fabulous advances in genetic technology currently taking place, it may become possible for a tyrant, terrorist, or lunatic to create a doomsday virus, an organism that combines long latency with high virulence and mortality…

Dangerous viruses can even be spawned unintentionally, as Australian researchers recently demonstrated when they created a modified mousepox virus with 100% mortality while trying to design a contraceptive virus for mice for use in pest control… While this particular virus doesn't affect humans, it is suspected that an analogous alteration would increase the mortality of the human smallpox virus. What underscores the future hazard here is that the research was quickly published in the open scientific literature… It is hard to see how information generated in open biotech research programs could be contained no matter how grave the potential danger that it poses; and the same holds for research in nanotechnology.

Genetic medicine will also lead to better cures and vaccines, but there is no guarantee that defense will always keep pace with offense. (Even the accidentally created mousepox virus had a 50% mortality rate on vaccinated mice.) Eventually, worry about biological weapons may be put to rest through the development of nanomedicine, but while nanotechnology has enormous long-term potential for medicine… it carries its own hazards.

Accidental misuse of nanotechnology ("gray goo")
The possibility of accidents can never be completely ruled out. However, there are many ways of making sure, through responsible engineering practices, that species-destroying accidents do not occur. One could avoid using self-replication; one could make nanobots dependent on some rare feedstock chemical that doesn't exist in the wild; one could confine them to sealed environments; one could design them in such a way that any mutation was overwhelmingly likely to cause a nanobot to completely cease to function… Accidental misuse is therefore a smaller concern than malicious misuse…

However, the distinction between the accidental and the deliberate can become blurred. While "in principle" it seems possible to make terminal nanotechnological accidents extremely improbable, the actual circumstances may not permit this ideal level of security to be realized. Compare nanotechnology with nuclear technology.

---

[8] Dear: To further explore the idea that some form of artificial intelligence might "take over", I'd recommend that you read the article by Eliezer Yudkowsky entitled "Artificial Intelligence as a Positive and Negative Factor in Global Risk" at http://www.singinst.org/upload/artificial-intelligence-risk.pdf.

From an engineering perspective, it is of course perfectly possible to use nuclear technology only for peaceful purposes such as nuclear reactors, which have a zero chance of destroying the whole planet. Yet in practice it may be very hard to avoid nuclear technology also being used to build nuclear weapons, leading to an arms race. With large nuclear arsenals on hair-trigger alert, there is inevitably a significant risk of accidental war. The same can happen with nanotechnology: it may be pressed into serving military objectives in a way that carries unavoidable risks of serious accidents.

In some situations it can even be strategically advantageous to *deliberately* make one's technology or control systems risky, for example in order to make a "threat that leaves something to chance."

Something unforeseen
We need a catch-all category. It would be foolish to be confident that we have already imagined and anticipated all significant risks. Future technological or scientific developments may very well reveal novel ways of destroying the world.[9]

Physics disasters
The Manhattan Project bomb-builders' concern about an A-bomb-derived atmospheric conflagration has contemporary analogues.

There have been speculations that future high-energy particle accelerator experiments may cause a breakdown of a metastable vacuum state that our part of cosmos might be in, converting it into a "true" vacuum of lower energy density… This would result in an expanding bubble of total destruction that would sweep through the galaxy and beyond at the speed of light, tearing all matter apart as it proceeds.

Another conceivability is that accelerator experiments might produce negatively charged stable "strangelets" (a hypothetical form of nuclear matter) or create a mini black hole that would sink to the center of the Earth and start accreting the rest of the planet…

These outcomes *seem* to be impossible given our best current physical theories. But of course, the reason why we do the experiments is precisely that we don't really

---

[9] Some *foreseen* hazards which have been excluded from the list of Bangs on grounds that they seem too unlikely to cause a global terminal disaster are: solar flares, supernovae, black hole explosions or mergers, gamma-ray bursts, galactic center outbursts, supervolcanos, loss of biodiversity, buildup of air pollution, gradual loss of human fertility, and various religious doomsday scenarios. The hypothesis that we will one day become "illuminated" and commit collective suicide or stop reproducing, as supporters of VHEMT (The Voluntary Human Extinction Movement) hope… appears unlikely. If it really were better not to exist (as Silenus told king Midas in the Greek myth, and as Arthur Schopenhauer argued… although for reasons specific to his philosophical system he didn't advocate suicide), then we should not count this scenario as an existential disaster. The assumption that it is not worse to be alive should be regarded as an implicit assumption in the definition of Bangs. *Erroneous* collective suicide is an existential risk albeit one whose probability seems extremely slight…

know what will happen. A more reassuring argument is that the energy densities attained in present day accelerators are far lower than those that occur naturally in collations between cosmic rays… It's possible, however, that factors other than energy density are relevant for these hypothetical process, and that those factors will be brought together in novel ways in future experiments.

The main reason for concern in the "physics disasters" category is the meta-level observation that discoveries of all sorts of weird physical phenomena are made all the time, so even if right now all the particular physics disasters we have conceived of were absurdly improbable or impossible, there could be other more realistic failure-modes waiting to be uncovered. The ones listed here are merely illustrations of the general case. [For example, Dear, you might want to check out the recent discovery of a Black Hole at the center of one of a pair of galaxies (3C321) that is irradiating a part of the other galaxy, probably eliminating essentially all life (if it exists) on any planets (if they exist) encircling millions of stars – and then remember that the Andromeda galaxy is headed towards joining up with our Milky Way Galaxy![10]]

Naturally occurring disease
What if AIDS had been as contagious as the common cold?

There are several features of today's world that may make a global pandemic more likely than ever before. Travel, food-trade, and urban dwelling have all increased dramatically in modern times, making it easier for a new disease to quickly infect a large fraction of the world's population.

Asteroid or comet impact
There is a real but very small risk that we will be wiped out by the impact of an asteroid or comet…

In order to cause the extinction of human life, the impacting body would probably have to be greater than 1 km in diameter (and probably 3 - 10 km). There have been at least five and maybe well over a dozen mass extinctions on Earth, and it is quite probable that at least some of these have been caused by impacts… In particular, the K/T extinction 65 million years ago, in which the dinosaurs went extinct, has been linked to the impact of an asteroid between 10 and 15 km in diameter on the Yucatan peninsula. It is estimated that a 1 km or greater body collides with Earth about once every 0.5 million years.[11] We have only catalogued a small fraction of the potentially hazardous bodies.

---

[10] For example, check out http://www.timesonline.co.uk/tol/news/uk/science/article3065514.ece or http://www.nasa.gov/home/hqnews/2007/dec/HQ_07280_Death_Star_Black_Hole.html.

[11] By comparison, the Tunguska event in 1908 was caused by a body about 60 meters in diameter, producing a yield of 2 megatons TNT (the Hiroshima bomb had a yield of 2 kilotons) and felling trees within a 40 km radius.

If we were to detect an approaching body in time, we would have a good chance of diverting it by intercepting it with a rocket loaded with a nuclear bomb…

Runaway global warming
One scenario is that the release of greenhouse gases into the atmosphere turns out to be a strongly self-reinforcing feedback process. Maybe this is what happened on Venus, which now has an atmosphere dense with CO_2 and a temperature of about 450° C. Hopefully, however, we will have technological means of counteracting such a trend by the time it would start getting truly dangerous.

Crunches
While some of the events described in the previous section would be certain to actually wipe out Homo sapiens (e.g. a breakdown of a meta-stable vacuum state) others could potentially be survived (such as an all-out nuclear war). If modern civilization were to collapse, it is not completely certain that it would arise again, however, even if the human species survived. We *may* have used up too many of the easily available resources that a primitive society would need in order to use to work itself up to our level of technology. A primitive human society may or may not be more likely to face extinction than any other animal species. But let's not try that experiment.

If the primitive society lives on but fails to ever get back to current technological levels, let along go beyond it, then we have an example of a crunch. Here are some potential causes of a crunch:

Resource depletion or ecological destruction
The natural resources needed to sustain a high-tech civilization are being used up. If some other cataclysm destroys the technology we have, it may not be possible to climb back up to present levels if natural conditions are less favorable than they were for our ancestors, for example if the most easily exploitable coal, oil, and mineral resources have been depleted. (On the other hand, if plenty of information about our technological feats is preserved, that could make a rebirth of civilization easier.)

Misguided world government decides to stop technological progress
One could imagine that some fundamentalist religious or ecological movement [may] one day come to dominate the world. If by that time there are means of making such a world government stable against insurrections (by advanced surveillance or mind-control technologies), this might permanently put a lid on humanity's potential to develop to a posthuman level. Aldous Huxley's *Brave New World* is a well-known scenario of this type…

"Dysgenic" pressures
It is possible that advanced civilized society is dependent on there being a sufficiently large fraction of intellectually talented individuals. Currently it seems that there is a negative correlation in some places between intellectual achievement and fertility. If

such selection were to operate over a long period of time, we might evolve into a less brainy but more fertile species, *homo philoprogenitus* ("lover of many offspring").

However, contrary to what such considerations might lead one to suspect, IQ scores have actually been increasing dramatically over the past century. This is known as the Flynn effect… It's not yet settled whether this corresponds to real gains in important intellectual functions.

Moreover, genetic engineering is rapidly approaching the point where it becomes possible to give parents the choice of endowing their offspring with genes that correlate with intellectual capacity, physical health, longevity, and other desirable traits.

In any case, the time-scale for human natural genetic evolution seems much too grand for such developments to have any significant effect before other developments will have made the issue moot…

Technological arrest
The shear technological difficulties in making the transition to the posthuman world might turn out to be so great that we never get there.

Something unforeseen[12]
Overall, the probability of a crunch seems much smaller than that of a bang. We should bear the possibility in mind but not let it play a dominant role in our thinking at this point. If technological and economical development for some reason were to slow down substantially, then we should have to take a closer look at the crunch scenarios.

Shrieks
Determining what scenarios are shrieks is made more difficult by the inclusion of the notion of *desirability* in the definition. Unless we know what is "desirable", we cannot tell what scenarios are shrieks. However, there are some scenarios that would count as shrieks on most reasonable interpretations.

Take-over by a transcending upload
Suppose uploads come before human-level artificial intelligence. An upload is a mind that has been transferred from biological brain to a computer that emulates the computational processes that took place in the original biological neural network… A successful uploading process would preserve the original mind's memories, skills, values, and consciousness. Uploading a mind will make it much easier to enhance its intelligence, by running it faster, adding additional computational resources, or streamlining its architecture. One could imagine that enhancing an upload beyond a certain point will result in a positive feedback loop, where the enhanced upload is

[12] It is questionable whether a badly programmed superintelligence that decided to hold humanity back indefinitely could count as a whimper. The superintelligence would have to be of such a limited nature that it wouldn't itself count as some form of posthumanity; otherwise this would be a shriek.

able to figure out ways of making itself even smarter; and the smarter successor version is in turn even better at designing an improved version of itself, and so on. If this runaway process is sudden, it could result in one upload reaching superhuman levels of intelligence while everybody else remain at a roughly human level. Such enormous intellectual superiority may well give it a correspondingly great power. It could rapidly invent new technologies or perfect nanotechnological designs, for example. If the transcending upload is bent on preventing others from getting the opportunity to upload, it might do so.

The posthuman world would then be a reflection of one particular egoistical upload's preferences (which in a worst case scenario would be worse than worthless). Such a world may well be a realization of only a tiny part of what would have been possible and desirable. This end is a shriek.

Flawed superintelligence
Again, there is the possibility that a badly programmed superintelligence takes over and implements the mistaken goals it has been given.

Repressive totalitarian global regime
Similarly, one can imagine that an intolerant world government, based perhaps on mistaken religious or ethical convictions, is formed, is stable, and decides to realize only a very small part of all the good things a posthuman world could contain.

Such a world government could conceivably be formed by a small group of people if they were in control of the first superintelligence and could select its goals. If the superintelligence arises suddenly and becomes powerful enough to take over the world, the posthuman world may reflect only the idiosyncratic values of the owners or designers of this superintelligence. Depending on what those values are, this scenario would count as a shriek.

Something unforeseen.[13]
These shriek scenarios appear to have substantial probability and thus should be taken seriously in our strategic planning.

One could argue that one value that makes up a large portion of what we would consider desirable in a posthuman world is that it contains as many as possible of those persons who are currently alive. After all, many of us want very much not to die (at least not yet) and to have the chance of becoming posthumans. If we accept this, then *any* scenario in which the transition to the posthuman world is delayed for long enough that almost all current humans are dead before it happens (assuming they have not been successfully preserved via cryonics arrangements…) would be a shriek. Failing a breakthrough in life-extension or widespread adoption of cryonics, then even a smooth transition to a fully developed posthuman eighty years from now

[13] I regard the hypothesis (common in the mass media…) that we will be exterminated in a conventional war between the human species and a population of roughly human-equivalent human-made robots as extremely small.

would constitute a major existential risk, *if* we define "desirable" with special reference to the people who are currently alive. This "if", however, is loaded with a profound axiological [viz., "values"] problem that we shall not try to resolve here.

Whimpers

If things go well, we may one day run up against fundamental physical limits. Even though the universe appears to be infinite… the portion of the universe that we could potentially colonize is (given our current admittedly very imperfect understanding of the situation) finite… and we will therefore eventually exhaust all available resources or the resources will spontaneously decay through the gradual decrease of negentropy [or increase in entropy] and the associated decay of matter into radiation. But here we are talking astronomical time-scales. An ending of this sort may indeed be the best we can hope for, so it would be misleading to count it as an existential risk. It does not qualify as a whimper because humanity could on this scenario have realized a good part of its potential.

Two whimpers (apart form the usual catch-all hypothesis) appear to have significant probability:

Our potential or even our core values are eroded by evolutionary development

This scenario is conceptually more complicated than the other existential risks we have considered (together perhaps with the "We are living in a simulation that gets shut down" bang scenario). It is explored in more detail in a companion paper [by the author]. An outline of that paper is provided in an Appendix [which, Dear, I've added here:

1. Although it's easy to think of evolution as leading from simple to more complex life forms, we should not uncritically assume that this is always so. It is true that here on Earth, simple replicators have evolved to human beings (among other things), but because of an observational selection effect the evidential value of this single datapoint is very limited…

2. We don't currently *see* much evolutionary development in the human species. This is because biological evolution operates on a time-scale of many generations, not because it doesn't occur any longer…

3. Biological human evolution is slow primarily because of the slowness of human reproduction (with a minimum generational lag of about one and a half decade).

4. Uploads and machine intelligences can reproduce virtually instantaneously provided easy resources are available. Also, if they [could] predict some aspects of their evolution, they [could] modify themselves accordingly now rather than waiting to be outcompeted. Both these factors can lead to a much more rapid evolutionary development in a posthuman world.

5. The activities and ways of being that we attach value to may not coincide with the activities that have the highest economic value in the posthuman world. Agents who choose to devote some fraction of their resources to "hobbies" would be at a competitive disadvantage, and would therefore risk being outcompeted.

6. We need to distinguish between two senses of "outcompeted": In the first sense, an outcompeted type would possess a smaller and smaller fraction of all colonized resources; in

the second sense, its possessions would decrease in absolute terms and the type would eventually become extinct.

7. If property rights are nearly perfectly enforced (over cosmic distances, which seems hard to do) then the "hobbyists" would be outcompeted only in the first sense. Depending on the details, this may or may not qualify as a whimper. If the lost potential (due to the increasing dominance of types that we don't regard as valuable) is great enough, it would be a whimper.

8. Without nearly perfect enforcement of property rights, we would have to fear that the hobbyists would become extinct because they are less efficient competitors for the same ecological niche as those types which don't expend any of their resources on hobbyist activities.

9. The only way of avoiding this outcome may be to replace natural evolution with *directed evolution*, by shaping the social selection pressures so that they favor the hobbyist type (for example by taxing the non-hobbyists)…

10. Directed evolution, however, requires coordination. It is no good if some societies decide to favor their hobbyists if there are other societies that instead decide to maximize their productivity. For the latter would then eventually outcompete the former. Therefore, on the highest level there must be only one agent – a singleton – if directed evolution is to succeed in avoiding ending with a whimper.

11. A singleton does not need to be a monolith. It can contain within itself a very diverse ecology of independent groups and individuals. A singleton could for example be a democratic world government or a friendly superintelligence…]

A related scenario… argues that our "cosmic commons" could be burnt up in a colonization race. Selection would favor those replicators that spend *all* their resources on sending out further colonization probes… Although the time it would take for whimper of this kind to play itself out may be relatively long, it could still have important policy implications because near-term choices may determine whether we will go down a track… that inevitably leads to this outcome. Once the evolutionary process is set in motion or a cosmic colonization race begun, it could prove difficult or impossible to halt it… It may well be that the only feasible way of avoiding a whimper is to prevent these chains of events from ever starting to unwind.

Killed by an extraterrestrial civilization

The probability of running into aliens any time soon appears to be very small… If things go well, however, and we develop into an intergalactic civilization, we may one day in the distant future encounter aliens. If they were hostile and if (for some unknown reason) they had significantly better technology than we will have then, they may begin the process of conquering us. Alternatively, if they trigger a phase transition of the vacuum through their high-energy physics experiments (see the Bangs section) we may one day face the consequences. Because the spatial extent of our civilization at that stage would likely be very large, the conquest or destruction would take relatively long to complete, making this scenario a whimper rather than a bang.

Something unforeseen
The first of these whimper scenarios should be a weighty concern when formulating long-term strategy. Dealing with the second whimper is something we can safely delegate to future generations (since there's nothing we could do about it now anyway).

Assessing the probability of existential risks
Direct versus indirect methods
There are two complementary ways of estimating our chances of creating a posthuman world. What we could call the *direct way* is to analyze the various specific failure-modes, assign them probabilities, and then subtract the sum of these disaster-probabilities from unity to get the success-probability. In doing so, we would benefit from a detailed understanding of how the underlying causal factors will play out. For example, we would like to know the answers to questions such as: How much harder is it to design a foolproof global nanotech immune system than it is to design a nanobot that can survive and reproduce in the natural environment? How feasible is it to keep nanotechnology strictly regulated for a lengthy period of time (so that nobody with malicious intentions get their hands on an assembler that is not contained in a tamperproof sealed assembler lab…)? How likely is it that superintelligence will come before advanced nanotechnology? We can make guesses about these and other relevant parameters and form an estimate on the basis of that; and we can do the same for the other existential risks that we have outlined above. (I have tried to indicate roughly the relative probability of the various risks in the rankings given in the previous four sections.)

Secondly, there is the *indirect way*. There are theoretical constraints that can be brought to bear on the issue, based on some general features of the world we are living in. There is only small number of these, but they are important because they do not rely on making a lot of guesses about the details of future technological and social developments:

The Fermi Paradox
The Fermi Paradox refers to the question mark that hovers over the data point that we have seen no signs of extraterrestrial life… This tells us that it is not the case that life evolves on a significant fraction of Earth-like planets and proceeds to develop advanced technology, using it to colonize the universe in ways that would have been detected with our current instrumentation. There must be (at least) one Great Filter – an evolutionary step that is extremely improbable – somewhere on the line between Earth-like planet and colonizing-in-detectable-ways civilization… If the Great Filter isn't in our past, we must fear it in our (near) future. Maybe nearly every civilization that develops a certain level of technology causes its own extinction.

Luckily, what we know about our evolutionary past is consistent with the hypothesis that the Great Filter is behind us. There are several plausible candidates for evolutionary steps that may be sufficiently improbable to explain why we haven't seen or met any extraterrestrials, including: the emergence of the first organic self-

replicators; the transition from prokaryotes to eukaryotes; to oxygen breathing; and to sexual reproduction; and possibly others.[14] The upshot is that at the current state of the art in evolutionary biology, Great Filter arguments cannot tell us much about how likely we are to become posthuman, although they may give us subtle hints…

This would change dramatically if we discovered traces of life (whether extinct or not) on other planets. Such a discovery would be bad news. Finding a relatively advanced life-form (multicellular organisms) would be especially depressing.

Observational selection effects

The theory of observational selection effects may tell us what we should expect to observe given some hypothesis about the distribution of observers in the world. By comparing such predictions to our actual observations, we get probabilistic evidence for or against various hypotheses.

One attempt to apply such reasoning to predicting our future prospects is the so-called Doomsday argument… It purports to show that we have systematically underestimated the probability that humankind will go extinct relatively soon. The idea, in its simplest form, is that it is we [who] should think of ourselves as in some sense random samples from the set of all observers in our reference class, and we would be more likely to live as early as we do if there were not a very great number of observers in our reference class living later than us. The Doomsday argument is highly controversial, and I have argued elsewhere that although it may be theoretically sound, some of its applicability conditions are in fact not satisfied, so that applying it to our actual case would be a mistake…

Other anthropic arguments may be more successful: the argument based on the Fermi-paradox is one example and the next section provides another. In general, one lesson is that we should be careful not to use the fact that life on Earth has survived up to this day and that our humanoid ancestors didn't go extinct in some sudden disaster to infer the thesis that that Earth-bound life and humanoid ancestors are highly resilient. Even if on the vast majority of Earth-like planets life goes extinct before intelligent life forms evolve, we should still expect to find ourselves on one of

---

[14] These are plausible candidates for difficult, critical steps (perhaps requiring simultaneous muli-loci mutations or other rare coincidences) primarily because they took a very long time (by contrast, for instance, of the evolution of Homo sapiens sapiens from our humanoid ancestors). [Dear: the extra 'sapiens', there, isn't a misprint; it refers to the "subspecies of Homo sapiens that includes all modern 'races'."] Yet the duration of a step is not always good reason for thinking the step improbable. For example, oxygen breathing took a long time to evolve, but this is not a ground for thinking that it was a difficult step. Oxygen breathing became adaptive only after there were significant levels of free oxygen in the atmosphere, and it took anaerobic organisms hundreds of millions of years to produce enough oxygen to satiate various oxygen sinks and raise the levels of atmospheric oxygen to the required levels. This process was very slow but virtually guaranteed to run to completion eventually, so it would be a mistake to infer that the evolution of oxygen breathing and the concomitant Cambrian explosion represent a hugely difficult step in human evolution.

the exceptional planets that were lucky enough to escape devastation.[15] In this case, our past success provides no ground for expecting success in the future.

The field of observational selection effects is methodologically very complex… and more foundational work is needed before we can be confident that we really understand how to reason about these things. There may well be further lessons from this domain that we haven't yet been clever enough to comprehend.

The Simulation argument

Most people don't believe that they are currently living in a computer simulation. I've recently shown (using only some fairly uncontroversial parts of the theory of observational selection effects) that this commits one to the belief that either we are almost certain never to reach the posthuman stage or almost all posthuman civilizations lack individuals who run large numbers of ancestor-simulations – i.e., computer-emulations of the sort of human-like creatures from which they evolved… This conclusion is a pessimistic one, for it narrows down quite substantially the range of positive future scenarios that are tenable in light of the empirical information we now have.

The Simulation argument does more than just sound a general alarm; it also redistributes probability among the hypotheses that remain believable. It increases the probability that we are living in a simulation (and that may in many subtle ways affect our estimates of how likely various outcomes are) and it decreases the probability the posthuman world would contain lots of free individuals that have large resources and human-like motives. This gives us some valuable hints as to what we may realistically hope for and consequently where we should direct our efforts.

Psychological biases?

The psychology of risk perception is an active but rather messy … that could potentially contribute indirect grounds for reassessing our estimates of existential risks… Suppose that our intuitions about which future scenarios are "plausible and realistic" are shaped by what we see on TV and in movies and what we read in novels… We should then, when thinking critically, suspect our intuitions of being biased in the direction of overestimating the probability of those scenarios which make for a good story, since such scenarios will seem much more familiar and more "real". This *Good-story bias* could be quite powerful. When was the last time you saw a movie about humankind suddenly going extinct (without warning and without being replaced by some other civilization)? While this scenario may be much more probable than a scenario in which human heroes successfully repel an invasion of monsters or robot warriors, it wouldn't be much fun to watch. So we don't see many stories of that kind. If we are not careful, we can be misled into believing that the boring scenario is too farfetched to be worth taking seriously. In general, if we think there is a Good-story bias, we may want upon reflection to increase our credence in

[15] This holds so long as the total number of Earth-like planets in cosmos is sufficiently great to make it highly likely that at least some of them would develop intelligent observers…

boring hypotheses and decrease our credence in interesting, dramatic hypotheses. The net effect would be to redistribute probability among existential risks in favor of those that seem to [be] harder to fit into a selling narrative, and possibly to increase the probability of the existential risks as a group.

The empirical data on risk-estimation biases [are] ambiguous. It has been argued that we suffer from various systematic biases when estimating our own prospects or the risks in general. Some data suggest that humans tend to overestimate their own personal abilities and prospects.[16] About three quarters of all drivers think they are safer drivers than the typical driver.[17] Bias seems to be present even among highly educated people. According to one survey, almost half of all sociologists believed that they would become one of the top ten in their field… and 94% of sociologists thought they were better at their jobs than their average colleagues… It has also been shown that depressives have a more accurate self-perception than normals except regarding the hopelessness of their situation… Most people seem to think that they themselves are less likely to fall victims to common risks than other people… It is widely believed… that the public tends to overestimate the probability of highly publicized risks (such as plane crashes, murders, food poisonings etc.), and a recent study… shows the public overestimating a large range of commonplace health risks to themselves. Another recent study… however, suggests that available data [are] consistent with the assumption that the public rationally estimates risk (although with a slight truncation bias due to cognitive costs of keeping in mind exact information).[18] Even if we could get firm evidence for biases in estimating personal risks, we'd still have to be careful in making inferences to the case of existential risks.

Weighing up the evidence

In combination, these indirect arguments add important constraints to those we can glean from the direct consideration of various technological risks, although there is not room here to elaborate on the details. But the balance of evidence is such that it would appear unreasonable not to assign a substantial probability to the hypothesis that an existential disaster will do us in. My subjective opinion is that setting this probability lower than 25% would be misguided, and the best estimate may be considerably higher. But even if the probability were much smaller (say, ~1%) the subject matter would still merit very serious attention because of how much is at stake. [Italics added.]

---

[16] Or at least that males do. One review… suggests that females underestimate their prospects although not by as much as males overestimate theirs…

[17] …Some of these studies neglect that it may well be *true* that 75% of drivers are better than the average driver; some studies, however, seem to avoid this problem…

[18] Could the reason why recent studies speak more favorably about public rational risk assessment be that earlier results have resulted in public learning and recalibration? Researchers trying to establish systematic biases in risk perception could be shooting after a moving target much like those who attempt to find regularities in stock indexes. As soon as a consensus develops that there is such an effect, it disappears.

In general, the greatest existential risks on the time-scale of a couple of centuries or less appear to be those that derive from the activities of advanced technological civilizations. We see this by looking at the various existential risks we have listed. In each of the four categories, the top risks are engendered by our activities. The only significant existential risks for which this isn't true are "simulation gets shut down" (although on some versions of this hypothesis the shutdown would be prompted by our activities…); the catch-all hypotheses (which include both types of scenarios); asteroid or comet impact (which is a very low probability risk); and getting killed by an extraterrestrial civilization (which would be highly unlikely in the near future).[19]

It may not be surprising that existential risks created by modern civilization get the lion share of the probability. After all, we are now doing some things that have never been done on Earth before, and we are developing capacities to do many more such things. If non-anthropogenic factors have failed to annihilate the human species for hundreds of thousands of years, it could seem unlikely that such factors will strike us down in the next century or two. By contrast, we have no reason whatever not to think that the products of advanced civilization will be our bane.

We shouldn't be too quick to dismiss the existential risks that aren't human-generated as insignificant, however. It's true that our species has survived for a long time in spite of whatever such risks are present. But there may be an observational selection effect in play here. The question to ask is, on the theory that natural disasters sterilize Earth-like planets with a high frequency, what should we expect to observe? Clearly not that we are living on a sterilized planet. But maybe that we should be more primitive humans than we are? In order to answer this question, we need a solution to the problem of the reference class in observer selection theory… But that is a part of the methodology that doesn't yet exist. So at the moment we can state that the most serious existential risks are generated by advanced human civilization but we base this assertion on direct considerations. Whether there is additional support for it based on indirect considerations is an open question.

We should not *blame* civilization or technology for imposing big existential risks. Because of the way we have defined existential risks, a failure to develop technological civilization would imply that we had fallen victims of an existential disaster (namely a crunch, "technological arrest"). Without technology, our chances of avoiding existential risks would therefore be nil. With technology, we have some chance, although the greatest risks now turn out to be those generated by technology itself.

Implications for policy and ethics

Existential risks have a cluster of features that make it useful to identify them as a special category: the extreme magnitude of the harm that would come from an existential disaster; the futility of the trial-and-error approach; the lack of evolved biological and cultural coping methods; the fact that existential risk dilution is a

[19] The crunch scenario "technological arrest" couldn't properly be said to be *caused* by our activities.

> global public good; the shared stakeholdership of all future generations; the international nature of many of the required countermeasures; the necessarily highly speculative and multidisciplinary nature of the topic; the subtle and diverse methodological problems involved in assessing the probability of existential risks; and the comparative neglect of the whole area. From our survey of the most important existential risks and their key attributes, we can extract tentative recommendations for ethics and policy:
>
> *Raise the profile of existential risks*
> We need more research into existential risks – detailed studies of particular aspects of specific risks as well as more general investigations of associated ethical, methodological, security and policy issues. Public awareness should also be built up so that constructive political debates about possible countermeasures become possible.
>
> Now, it's commonplace that researchers always conclude that more research needs to be done in their field. But in this instance it is *really* true. There is more scholarly work on the life-habits of the dung fly than on existential risks.
>
> *Create a framework for international action*
> Since existential risk reduction is a global public good, there should ideally be an institutional framework such that the cost and responsibility for providing such goods could be shared by fairly by all people. Or even if the costs couldn't be shared fairly, at least some system that leads to the provision of existential risk reduction in something approaching optimal amounts.

Bostrom's closing call for "an institutional framework… for international action" to assess "existential risks" has recently been echoed by many other scientists. An example is described in a 17 March 2012 article in *Science* (Vol. 335, p. 1306) by F. Biermann and 31 other authors entitled "Navigating the Anthropocene: Improving Earth System Governance". Their introductory paragraph is:

> Science assessments indicate that human activities are moving several of Earth's sub-systems outside the range of natural variability typical for the previous 500,000 years… Human societies must now change course and steer away from critical tipping points in the Earth system that might lead to rapid and irreversible change… This requires fundamental reorientation and restructuring of national and international institutions toward more effective Earth system governance and planetary stewardship.

To pursue additional details, Dear, you may want to start at the website of the Earth System Governance Project at http://earthsystemgovernance.org/.

My second long – and much more "upbeat" – quotation is from Clayton Naff. He presented it as a "sermon", as you can see in the heading of the following.[20]

How to Have Faith Without Superstition

Clayton Farris Naff

Adapted from a service given at the
First Unitarian Church of Lubbock
October 7, 2001 (rev. 12/01)

If the Bible is wrong about cosmology, if it's wrong about geology, if it's wrong about biology, why should I trust it about morality and salvation?
[Ken Ham, fundamentalist preacher and founder of Answers in Genesis]

Hope springs eternal in the human breast:
Man never is, but always to be, blest.
[Alexander Pope, *Essay on Man*]

We are a way for the Cosmos to know itself.
[Carl Sagan, scientist and humanist]

Why are we here? This is perhaps the deepest question ever asked, and in all humility I propose to offer a new answer. In doing so, it will be necessary to discredit both the traditional religious answers and the contemporary atheistic answers.

I attempt this not because I think I am the smartest person ever to consider the question (I'm not). It's just that I have the advantage of thinking about it at a special time in history (more on that later). Nor do I seek to bring down the old answers simply for my satisfaction. I do it because I sincerely believe it is a matter of human survival.

If Alexander Pope were here today, he might fling a couplet or two at me:

Presumptuous man! the reason wouldst thou find,
Why form'd so weak, so little, and so blind?

---

[20] Dear: Unfortunately, this "sermon" may no longer be available on the internet; recently, I couldn't find it. If you want to learn more about Naff, see http://www.zoominfo.com/people/Naff_Clay_4037620.aspx, where I learned that this "secular humanist" was a "journalist, researcher, and former director of Community Action of Nebraska" and is now "working on a book, tentatively called *Making Peace with God: How to Transform Religion Before All Hell Breaks Loose.*"

And indeed you may also wonder, who put me in charge of human fate? To those who have faith in a living God, the matter rests with him. To the fundamentalist Christian, the Rapture may start at any moment. To the hopeful Jew, the Messiah will eventually return. To the devout Muslim, Allah is ever watchful over the faithful. So I can look for no sympathy from religionists.

And I'm afraid I'll fare no better with the naturalists. In the minds of most who have a strictly scientific outlook, our destiny is set. We must in the end share the fate of our universe, a long unwinding that ends in the quiet twitch of lonely particles.

Who is right? In these postmodern days, it is considered not just bad form but practically a form of oppression to say that anyone's worldview might be wrong. I'm well aware of this. But I will say. Both science and religion are wrong.

How can I make such a sweeping assertion? Consider. Even if we accept that there is no absolute certainty in this world, that there is only a scale of veridicality, a continuum of credibility, a range of probability, even if we accept all this, we must have some way of sorting out what is likely true from what is likely false.

Fortunately, we do have a way. It is called reason. And when we harness reason to systematic investigations of the evidence out there, we can reach conclusions that, while always tentative, nevertheless deserve the name of certainty.

At the risk of offending some of you, I must confess: I am certain the story told by revealed religions is wrong. I cannot in the space of this short essay attempt to debunk every supernatural belief, so let me just say that, for me, Christianity, the dominant religion in our culture, the foundation stone of Unitarianism, and the faith of my ancestors, may be set aside with certainty.

Here is one reason why. The central prediction of Christianity is Christ's return. This is the key belief upon which believers stake their faith. Yet the Bible not only says he will return, it says he will return soon.

I am not accustomed to citing chapter and verse, but since this was originally written for a church service, let me share a bit of Scripture with you. First, a selection from *Revelations 22*, the testimony of "John," relating what the angel told him. This is what he says:

> Do not seal up the words of the prophecy of this book, because the time is near. Let him who does wrong continue to do wrong; let him who is vile continue to be vile; let him who does right continue to do right; and let him who is holy continue to be holy. Behold, I am coming soon! My reward is with me, and I will give to everyone according to what he has done. I am the Alpha and the Omega, the First and the Last, the Beginning and the End. (*Rev.* 22)

If you're attentive, you pick up some pretty clear indications in that passage of what "soon" means. But in *1 Corinthians 7,* St. Paul is even clearer. This is what he says:

> Because of the present crisis, I think that it is good for you to remain as you are. Are you married? Do not seek a divorce. Are you unmarried? Do not look for a wife. But if you do marry, you have not sinned; and if a virgin marries, she has not sinned. But those who marry will face many troubles in this life, and I want to spare you this.
>
> What I mean, brothers, is that the time is short. From now on those who have wives should live as if they had none; those who mourn, as if they did not; those who are happy, as if they were not; those who buy something, as if it were not theirs to keep; those who use the things of the world, as if not engrossed in them. For this world in its present form is passing away. (*1 Corinthians 7*)

Can anyone doubt that Paul thinks Christ is returning in a matter of weeks, if not days? But even if we make allowances, the Bible as a whole puts the lie to the part. Biblical chronology yields a fairly precise timeline. Following in the footsteps of Bishop Usher, Cambridge University's Vice Chancellor John Lightfoot, in the 19th century set the birth of Jesus at 4004 years after creation. And here we are, two-thousand and one years later.

You don't have to be a mathematician to see that, in the Biblical frame, one-third of all time – or at least, of all human time – has passed since the crucifixion of Christ. Not even the most generous, optimistic, broadminded and dare I say liberal interpretation can keep the word "soon" afloat upon such a vast ocean of time.

And so sinks the crucial prediction of the Second Coming, and with it all reason for me, at least, to invest faith, hope or credence in Christianity.

What of science, then? When we look into the future along the sightlines of nature, we come to a chilling conclusion. History is against us. More than 99 percent of all species that have ever lived are now extinct. Even if we're clever enough to beat those odds, the sun is running down, and will eventually die. But, like many of us, as it ages, the sun is getting fatter and, as it goes about its work, it is getting hotter. In about a billion years, it will be so close and so hot that the seas of earth will boil away, and with it all life. Even if we somehow escape that fate, there's worse to come.

The universe is expanding at an ever-growing rate. With each tick of the clock, it grows a little colder. With every stroke, it grows a little more disordered. In the long run, the very atoms of which we're composed may be unstable.

"Things," wrote William Butler Yeats, "fall apart. The Center cannot hold." In his poem, *The Second Coming,* the Irish Poet was writing of political upheaval. But Yeats may have been right about the stuff of nature. Things fall apart.

Listen to Harvard-trained astronomer Brian Schmidt, speaking on PBS last year. This is what he said.

> In the distant future there will be nothing in the universe left to see, there will just be us. And that seems to me to be the coldest, most horrible end that I could think of. It is… it is just… I don't know. It's creepy. (NOVA, November 21, 2000)

Insights such as this led Nobel physicist Steven Weinberg to his most notorious remark. This is what he said: "The more the universe seems comprehensible, the more it also seems pointless."

And that's the story science tells. No wonder so few people are willing to accept it as the last word. But if science is so right on the credibility continuum, how can I say it's wrong about our fate? That's easy. Science is wrong about everything. It's just less wrong than any other kind of knowledge we have. Isaac Asimov put it best in a brilliant essay whose title says it all: "The Relativity of Wrong." Asimov points out that science is an always-improving approximation of truth.

Writing to a young English major with postmodernist sensibilities, Asimov said:

> John, when people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if you think that thinking the earth is spherical is just as wrong as thinking the earth is flat, then your view is wronger than both of them put together.

Asimov sees scientific knowledge as a series of concentric rings around the truth. With each advance we get to a smaller, closer ring. I agree, nearly. Although the never-ending path of rational, evidence-bound discovery leads us on an inward spiral toward truth, it still has many twists and turns. The Deists of the 18th century thought the universe went like clockwork. But it turns out when you look at it up close, the universe is unimaginably weird and unpredictable. Particles pop in and out of existence, photons simultaneously travel every conceivable path to their destination. The fabric of space-time rolls up into 10 dimensions, tears and folds back in on itself. Very strange. But of all the weird, unpredictable things in world, the most difficult to anticipate, the wild card in the deck, is us.

Humanity is in fact the one thing we know of that could alter the fate of the universe. And so, we come to the heart of rational enlightenment. We have arrived at last at… DeoLogic [which, Dear, seems to mean "God logic"]. I told you at the outset I would offer an answer to that deepest question of all time, and you have waited patiently for me to do so. Let me therefore set aside poetry and literature and highfalutin rhetoric and get down to brass tacks.

Here is how DeoLogic takes the bare facts and builds a stairway to heaven:

We do not know how or where life began, but we know with great assurance that all life on earth is based on DNA. And DNA, you recall, is a double-stranded self-replicating molecule that can encode instructions for making a complete living creature, anything from a bacterium to a potted plant or parrot, or even a human being. But why? What's the point of showing off like that?

There's only one reason: DNA exists to perpetuate itself. Not by the action of any intelligence, so far as the evidence shows, but simply by default. All the DNA that didn't care about perpetuating itself is presumably long gone.

DNA imposes its one and only value on all its creations, including us. It does not give a toss whether any particular individual survives, suffers, or finds happiness. It only cares about life rafts of DNA continuing to flow down the river of time.

One of the strategies adopted by DNA is to create animals with vast executive discretion about how to accomplish the goal. That would be us.

In doing so, DNA has granted us minds. Now, we're obviously not the only animals with minds, but ours now have a special property. We're the only creature with the power to discover DNA and mess around with it. No other living thing on earth has the slightest idea that its body is full of double-helixed molecules telling it what to do.

This discovery, along with our other knowledge and abilities, frees us up to make just about any choice concerning our future. We could, if we cared to, choose voluntary self-extinction. Indeed, there is a web site dedicated to that very idea (http://www.vhemt.org/). [A website, Dear, that I'll quote from in a later chapter.] But that is not what DeoLogic affirms. From this foundation of facts, we begin to ascend the staircase.

If we accept DNA's dictate as the root of all our values, then we at last have a rational, nonarbitrary system of values. And this is what it says: *Life must endure. To endure among those who have a choice, life must be worth living. To be worth living, life must be pleasurable and purposeful.*

Here we have an ethos distinct from both religious morality and humanist ethics. If we extrapolate from this ethos, we can justly claim that the persistence of conscious life in the universe is the ultimate good. Then, at last, we have a clear mission for humanity.

Now, if we examine what it would take to perpetuate conscious life in the universe, it soon becomes apparent that that we will have to evolve toward something like a deity or Supreme Being. No easy task, I grant you. Among other challenges, we will have to overcome the most fundamental force in the Universe, the tendency toward decay. But I have strong reasons to believe we can do it.

Let me pause here to try to sum up what DeoLogic is all about. *It is the realization that we are not the children of God; we are the potential ancestors of God.* [Italics added.] And it is the faith that we can make that potential real. Let me be clear: when I speak of the future God, I mean an entirely natural being.

You may remember from biology class that DNA is short for deoxyribonucleic acid. Notice that the first three letters are "deo." Nothing mystical in that, but it is a happy coincidence. It encourages us to see the entire journey as an unbroken progression. And so the DeoLogic vision may be encapsulated in the phrase, "from DNA to Deity."

But how plausible is this, you may ask? After all, I'm no scientist. Perhaps this is yet another false hope, impossible in principle and unattainable in reality. Take comfort. It is at least possible. The distinguished physicist Freeman Dyson has gone to great effort to determine whether life can endure. His 137 equations on the subject were peer reviewed, found to be flawless, and published in the *Reviews of Modern Physics*. This is what he says:

> The universe grows constantly colder as it expands, and the supply of free energy is constantly dwindling. To many people this future of endless ice has seemed even more dismal than the future of cataclysmic fire. But the laws of cosmic ecology put these futures into a very different perspective. If the hypothesis of adaptability is correct, life has a clear preference for ice over fire. In an expanding universe, life can adapt itself as the eons go by, constantly matching its metabolism of energy to the falling temperature of its surroundings. Since we are assuming perfect adaptability, the rate of energy metabolism falls with the square of the temperature. This has the consequence that, in an expanding universe, life of any fixed degree of complexity can survive forever upon a finite store of energy. (*Infinite in All Directions,* p. 111)

All right, then. Let's accept that it's possible for perfectly adapted life to endure. How are we to achieve this? For the answers, I rely on physicist Michio Kaku, and his marvelous book, *Visions*. Kaku starts by putting us in our place with a thump. This is what he says:

> …we are a Type 0 civilization… We are like infants… Our civilization is so new that just a hundred years ago we still got most of our energy from burning wood and coal. (p. 324)

Kaku is borrowing from the Russian astronomer Nikolai Kardashev, who created a classification scheme for civilizations with cosmic ambitions. Our next rung on the ladder, and the most slippery step, is to go from a rag-tag collection of nations to being a planetary civilization. Then, and only then, will we be Type I. After this, we progress to interplanetary civilization, known as Type II, and from there… you guessed it, a galactic civilization known as Type III.

But how do we get up that first rung? For one thing, we need more and better government. Now, that may sound strange to even the most liberal gathering in the Panhandle of Texas. But I'm serious.

I don't need to remind you that we live in a world where terrible violence, injustice and exploitation persist. To hear many politicians rattle on, you might think that the answer is less regulation and more freedom. But free markets alone won't end child slavery, ecological destruction, or terrorism. On the other hand, government carries its own opportunities for violence, injustice and exploitation. And who takes advantage of these opportunities? People! The American genius has been to keep powermongers at bay through checks and balances. Yet, you can hardly call our system optimal.

What else can be done? We can call on God, or we can work toward worldwide democratic machine government. The twenty-first century version would be entirely different from what Boss Tweed and Richard Daley constructed in centuries gone by. Pictures that catch red-light-runners and automatically issue them tickets. Now, imagine an apparatus of order a million times more sophisticated. Imagine if we had a reliable apparatus that could identify lies, or better yet, determine truth. Already, DNA testing has freed many innocent people from jail. But the application of technology to government has only just begun. The trick – and it is no small trick – is to ensure it is applied in service of just and democratic government.

Ideally, machine government will operate in support of genuine democracy. It would be a world-girdling system of restraint on human – and corporate – excess. A filter, if you will, straining out cruelty, exploitation and destruction, while allowing the free flow of creativity, exploration and joy. Am I leaving out a few details? Well, maybe, but we must move on.

This is only the beginning. As humanity continues to develop its computing, communications and AI [Artificial Intelligence] capacity, it will create a meta-intelligence that hovers over us in much the way a benevolent god is imagined to act.

To grasp this, think of an ant colony. What is it? Simply a collection of ants. Yet the colony behaves in ways vastly more intelligent than any one ant could. In fact, you could justly compare ants and their colonies to people and society. Except for one thing. We're not ants. We make conscious decisions. If we set our minds to it, we can build a planetary civilization.

Imagine, now, that humanity has stabilized its damage to the environment, restrained its population growth, instituted a system of universal peace and justice. Next, it turns its eyes to the skies.

There, we will have to prepare to fend off eventual catastrophe. We may be able to do this by deflecting killer comets and asteroids, but we will also hedge our bets by sending out pioneers to other planets and dwelling places.

Indeed, we must do this eventually, if we are to survive. Earth, be she ever so lovely, cannot sustain us forever. Long before life's 5-billion-year lease is up, we'll have evolved into a Type III civilization. We'll be whizzing around the galaxy, setting up colonies wherever hospitable planets can be found.

Except it won't exactly be us. Long before, intelligent life will have undergone changes more dramatic than a caterpillar's transformation into a butterfly.

And it's all because of Einstein. You remember Albert Einstein. He's the man who showed us that everything's relative. If you went strolling off into the stars in a steadily accelerating spaceship, you could approach the speed of light within about a year. And at that speed, you could traverse the Milky Way in just 25 years. Unfortunately, though, when you got back you wouldn't be able to brag about it to your friends. For while you were out gallivanting near the speed of light, 2 million years would have passed for those of us left behind on Earth. Relativity. It's a time-consuming thing.

So to become a Type III civilization, intelligent life will have to jump platforms. We humans may go on living out our brief little lives, but Mind must get beyond these old-fashioned, short-lived brains and become something not only greater but much more enduring. I've already talked about a meta-intelligence. Now we must contemplate the emergence of a higher-order intelligence, with its own self-awareness and sense of purpose.

We can know almost nothing about this greater intelligence. But we can assume this: it must have the same mission, to complete the journey from DNA to deity. Remember, the only alternative is the extinction of life and the eventual death of the universe.

Once this jump happens, we – if we are still we – will be in the presence of something very much like what Moses met in the desert. Except that God, to give this entity a name, will not be a wrathful tribal god, nor even a supernatural being. It will just be life, doing its thing. Life, after all, is the self-organization of matter and energy into patterns that seek to preserve themselves. Only now we're contemplating patterns on a cosmic scale. Deo will be a purely natural consequence of life's struggle to beat the odds and break the seemingly unbreakable law: Things fall apart.

Of course, the Second Law of Thermodynamics may in the end prove unbreakable. But there is a loophole, an escape hatch for our hopes. It may be that an advanced civilization will be able to bring about the birth of a new universe, a fresh start, possibly to its own specifications. If this is so, then our natural deity will eventually emulate its ancestors and become the proud parent of a baby with unlimited potential.

> If we accept DeoLogic and aim for the stars, will we succeed? Who knows… On the one hand, we have the implacable forces of nature, including the dark side of human nature.
>
> Yet, on the other hand, we have the remarkable achievements of civilization. A thousand years ago our proudest accomplishments were not particularly distinct from what termites, ants and crickets do: build structures, go to war and make music. In just 100 years, the typical human has moved from the farm to the city, from reading by candlelight to watching TV by electric light. In my lifetime, we have become a spacefaring species and in the short span my children's lives, we have given birth to the Internet.
>
> *As I said at the outset, we live in a special time. It is special not only because we have discovered what DNA is up to and begun to learn how to tinker with it. This is a special time because humanity is engaged in a great conflict between the passionate desire to preserve traditional beliefs and ways, the tumultuous forces of techno-capitalism, and the dawning realization that only we can save ourselves.* [Italics added.]
>
> I call this the rapids of history. Most rapids, as you'll recall, lead to a plunge over the falls. But as the only species that has learned how to fly without any instructions from our DNA, I remain hopeful that we will yet soar.
>
> If we can only get through the rapids of history intact, if we can become a peaceful, planetary civilization intent on the mission of universal progress toward the Supreme Being, there is surely every reason to hope we will succeed.
>
> So, why are we here? This is what I say: we are here to prepare the way. We are here to enjoy being alive. We are here to become worthy ancestors of God.

For those of us who have yet to see any gods or extraterrestrials provide any help to humanity, it's easy to agree with Neff's comment (italicized, above) that "only we can save ourselves." For example, Dear, if you want to improve your health, then…