

Dinosaurs, dodos, humans?

There is, perhaps, a 50% chance that humankind will be annihilated this century, says **Nick Bostrom**. We should try to prevent this

Something like 99.9% of all species that ever lived are now extinct. When will our own species join the dinosaurs and the dodos? How could that happen? And what can we do to stave off the end?

An existential risk is one that threatens to annihilate Earth-originating intelligent life or permanently and drastically curtail its potential. Since we are still here, we know that no existential disaster has ever occurred. But, lacking experience of such disasters, we have not evolved mechanisms, biologically or culturally, for managing existential risks.

The human species has had long experience with hazards such as dangerous animals, hostile tribes and individuals, poisonous foods, automobile accidents, Chernobyl, Bhopal, volcanic eruptions, earthquakes, droughts, wars, epidemics of influenza, smallpox, black plague and AIDS. These types of disaster have occurred many times throughout history. Our attitudes towards risk have been shaped by trial and error as we have been trying to cope with such risks. Yet – tragic as those events are to the people immediately affected – they have not determined the long-term fate of our species. Even the worst of those catastrophes were mere ripples on the surface of the great sea of life.

Existential risks are a relatively novel phenomenon. With the exception of a species-destroying comet or asteroid impact (an extremely rare occurrence), there were probably no significant existential risks in human history until the mid-20th century and certainly none that it was within our power to do anything about.

Perception is reality

The first man-made existential risk may have been the inaugural detonation of the atomic bomb. At the time, there was some concern that the explosion might start a runaway chain reaction by “igniting” the atmosphere. Although we now know that such an outcome is physically impossible, an existential risk was present then. 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 possibility of a disaster. If we do not know whether something

is objectively risky or not, then it is risky in the subjective sense. The subjective sense is what we must base our decisions on.

The human species has survived natural hazards for hundreds of thousands of years. It is unlikely, then, that such a hazard should strike us down within the next century. By contrast, the activities of our own species are introducing completely novel possibilities for disaster. The most serious existential risks for humanity in the 21st century are of our own making. More specifically, they are related to anticipated technological developments.

Man-made danger

In this century, we may learn to create designer pathogens that combine extremely high virulence with easy transmissibility, long incubation time and resistance to medication. Physicists will be colliding elementary particles and creating new kinds of matter in accelerators at increasingly high energies, and unexpected and potentially catastrophic things could happen when they do this. Nuclear war remains a big concern. Although current stockpiles are probably insufficient to cause human extinction even in an all-out war, future arms races might lead to the build-up of much larger arsenals.

Molecular nanotechnology will, in its mature form, give us an unprecedented ability to con-

With the exception of a species-destroying comet or asteroid impact, there were probably no significant existential risks in human history until the mid-20th century and certainly none that it was within our power to do anything about



Large asteroids hit the earth every 500,000 years

trol the structure of matter and to create powerful new weapons systems. Or what if we built computers more intelligent than humans and with the ability to self-improve? Furthermore, we must take into account the possibility that new kinds of risk may emerge that are as unforeseeable as the hydrogen bomb was in 1905.

Maybe the world will end not with a bang but a whimper. In addition to sudden, cataclysmic disasters, there are various more gradual and subtle ways in which an existential catastrophe could occur. One scenario is a totalitarian world government that oppresses its citizens and bans the use of technology to enhance human capacities, depriving humanity of its potential. Such a government might use ubiquitous surveillance and new kinds of mind-control technologies to prevent reform or rebellion.

Alternatively, evolutionary developments could take us in an undesirable direction, eliminating traits central to human values, such as consciousness, intelligence or, even, playfulness. There is no biological law that evolution always leads to an increase in valuable attributes. While human evolution operates over very long timescales, technologically-assisted evolution could be much faster and evolution among a population of machine intelligences or human uploads could be extremely rapid. The wrong kind of cultural “evolution” could lead to stagnation and thorough debasement of human life.

Our planet is just a small part of a potential-

ly much bigger drama. Earth is but a tiny speck in a cosmos that contains billions of suns that are illuminating empty rooms. Every second, more energy and matter go to waste in our galaxy than the human species has used up throughout its existence. If things go well, we will one day learn to use some of these resources to create wonderful new civilizations and novel forms of life.

Invasion of the robotic probes

Some studies suggest that an uncoordinated colonization race could, instead, lead to the evolution of self-replicating robotic colonization probes that would use up these cosmic commons to no other purpose than to make more copies of themselves. An astronomical potential for valuable development would have been lost.

The topic of existential risks has been remarkably neglected, with few systematic studies. Canadian philosopher John Leslie, after reviewing the evidence, concluded that the risk of human extinction within the present century is 50%. Britain’s Astronomer Royal, Sir Martin Rees, also puts the risk at 50%. Richard Posner, an American judge and legal scholar, did not give a numerical estimate but concluded in his book on the subject that the risk is “serious and substantial”. I argued in a paper in 2001 that it would be misguided to set the probability at less than 25%.

Attempts to quantify existential risk inevitably involve a large helping of subjective judgment. And, of course, there may be a publication bias in that those who believe that the

risk is larger are more likely to publish books. Nevertheless, everybody who has seriously looked at the issue agrees that the risks are considerable. Even if the probability of extinction were merely 5%, or 1%, it would still be worth taking seriously in view of how much is at stake.

It is sad that humanity as a whole has not invested even a few million dollars to improve its thinking about how it may best ensure its own survival. Some existential risks are difficult to study in a rigorous way but we will not know what insights we might develop until we do the research. There are also some sub-species of existential risk that can be measured, such as the risk of a species-destroying meteor or asteroid impact. This particular risk turns out to be very small. A meteor or an asteroid would have to be considerably larger than 1km in diameter to pose an existential risk. Fortunately, such objects hit the Earth less than once in 500,000 years on average.

The magnitude of existential risks is not a fixed quantity – it becomes larger or smaller depending on human action. We can take deliberate steps to reduce many existential risks. For instance, Nasa, America’s space agency, is mapping out asteroids larger than 1km in its Spaceguard Survey. If we were to get sufficiently early warning of a looming impact, it might be possible to launch a rocket with a nuclear charge to deflect the juggernaut.

Some of the studies and countermeasures that would reduce existential risk would also be relevant for mitigating lesser hazards. A global, catastrophic risk is one that could cause tens or hundreds of millions of deaths even as it falls far short of terminating the human race. The same programme that monitors doomsday asteroids can also detect somewhat smaller objects that pose “merely” global catastrophic risks. Countermeasures against future designer pathogens might also be useful in combating naturally-occurring pandemics. Research and measures to reduce the chance of catastrophic, runaway global warming would also be useful for understanding and counteracting the much more plausible mid-range scenarios.

A great leader accepts responsibility for the long-term consequences of the policies he or she pursues. With regard to existential risks, the challenge is to neither ignore them nor to indulge in gloomy despondency but to seek understanding and to take the most cost-effective steps to make the world safer. **GA**



CV NICK BOSTROM

Nick Bostrom is director of the Future of Humanity Institute at the University of Oxford. He has a background in physics, philosophy, computational neuroscience and artificial intelligence.