Sustainable Humanity, Sustainable Nature: Our Responsibility
Pontifical Academy of Sciences, Extra Series 41, Vatican City 2014
Pontifical Academy of Social Sciences, Acta 19, Vatican City 2014
www.pas.va/content/dam/accademia/pdf/es41/es41-rees.pdf

Existential Risks

Martin Rees

In 1902, the young H.G. Wells gave a celebrated lecture at the Royal Institution in London. He spoke mainly in visionary mode. "Humanity", he said, "has come some way, and the distance we have travelled gives us some earnest of the way we have to go. All the past is but the beginning of a beginning; all that the human mind has accomplished is but the dream before the awakening". His rather purple prose still resonates more than a hundred years later – he realised that we humans aren't the culmination of emergent life.

But Wells wasn't an optimist. He also highlighted the risk of global disaster: "It is impossible to show why certain things should not utterly destroy and end the human story ... and make all our efforts vain ... something from space, or pestilence, or some great disease of the atmosphere, some trailing cometary poison, some great emanation of vapour from the interior of the Earth, or new animals to prey on us, or some drug or wrecking madness in the mind of man".

The concept of devastating threats to human survival is certainly not new. Millennialism is perhaps as old as civilisation itself, and a scientific awareness of cataclysmic natural risk – from volcanoes and asteroid impact – dates back many decades. Were Wells writing today, he would have been elated by the amazing advances of science, but even more anxious about its 'downside'. Revolutionary new technologies might lead – perhaps accidentally, and perhaps very rapidly, once a certain point is reached – to a catastrophe. Some of the risks that have been envisaged may indeed be science fiction; but others may be disquietingly real. In future decades, events with low probability but catastrophic consequences may loom high on the political agenda.

Over nearly all of Earth's history, the greatest threats to humanity have come from nature – disease, earthquakes, floods, asteroids and so forth. But now our perspective is very different. More worrying by far are hazards that we ourselves engender – those caused by a rising population of humans, all empowered by advancing technology and more demanding of resources. Humans now utilise 40 percent of the world's biomass; we are collectively affecting the world's climate and ravaging the biosphere. The Earth has existed for 45 million centuries, but this is the first when one species – ours – has the planet's future in its hands. We're deep into Paul Crutzen's 'anthropocene era'.

Technological global hazards

Many speakers at this conference have addressed the threats stemming from the collective impact of the world's growing population on the biosphere and climate. I shall therefore focus on a different category of threats: those that could be caused by rather few individuals via misuse – by error or by design – of ever more powerful technologies.

At any time in the Cold War era – when armament levels escalated beyond all reason – the superpowers could have stumbled towards Armageddon through muddle and miscalculation. During the days of the Cuba crisis I and my fellow-students participated anxiously in vigils and demonstrations. But we would have been even more scared had we then realised just how close we were to catastrophe. Kennedy was later quoted as saying that during the Cuba crisis the probability of war was "between one in three and evens". And only when he was long retired did Kennedy's defence secretary, Robert McNamara, state frankly that "[w]e came within a hairbreadth of nuclear war without realizing it. It's no credit to us that we escaped – Khrushchev and Kennedy were lucky as well as wise". Be that as it may, we were surely at far greater hazard from nuclear catastrophe than from anything nature could do. Europe and North America would have been devastated; and the rest of the world would have suffered a 'nuclear winter'. We'd likely have a million years' grace before suffering a *natural* disaster – a giant asteroid impact or volcano supereruption – that was as disastrous.

The threat of global devastation involving tens of thousands of H-bombs is thankfully in abeyance; there is, though, now more reason to worry that smaller nuclear arsenals might be used in a regional context, or even by terrorists. But we can't rule out, later in the century, a geopolitical realignment leading to a standoff between new superpowers. So a new generation may face its own 'Cuba' – and one that could be handled less well or less luckily than the 1962 crisis was.

But the thermonuclear threat is not the only one – and maybe now not the most serious one – that humans confront as a downside of advancing technology. The H-bomb stemmed from 20th century science. But we should now be even more anxious about the powerful 21st century technologies on which our civilisation increasingly depends.

There are benefits from living in an interconnected world. But in consequence we are increasingly dependent on elaborate networks: electric-power grids, air traffic control, international finance, just-in-time delivery, globally-dispersed manufacturing, and so forth. Unless these globalised networks are highly resilient, their manifest benefits could be outweighed by catastrophic (albeit rare) breakdowns – real-world analogues of what hap-

pened in 2008 to the financial system. Our great cities would be quickly paralysed without electricity. Supermarket shelves would be empty within a couple of days if supply chains were disrupted. Air travel can spread a pandemic worldwide within days, causing the gravest havoc in the shambolic but burgeoning megacities of the developing world. And social media can spread panic and rumour, and psychic and economic contagion, literally at the speed of light.

Because technology gives powerful leverage to small groups, or even individuals, we're vulnerable not just to accidental malfunctions that cascade globally, but to maliciously-triggered events that could have catastrophic consequences. IT and biotech have a dark side: they will present new threats more diverse and more intractable than nuclear weapons did.

The techniques and expertise for cyber attacks are becoming accessible to millions – they don't require large special purpose facilities like nuclear weapons. Cyber-sabotage efforts like 'Stuxnet', and frequent hacking of financial institutions, have already pushed these concerns up the political agenda. A recent report from the Pentagon's Science Board claimed that the impact of the most sophisticated (state-engineered) cyber-attack could be catastrophic enough to justify a nuclear response.

And, before too long, millions will have the capability and resources to misuse biotech, just as they can misuse cybertech today. Advances in synthetic biology offer huge potential for medicine and agriculture – but they amplify the risk of bioerror or bioterror. Just last year some researchers who'd shown that it was surprisingly easy to make an influenza virus both virulent and transmissible were pressured to redact some details of their publication. The concern here was partly that it would be aiding terrorists, but partly also that if such experiments weren't conducted everywhere to the very highest safety and containment standards, there would be a risk of bioerror.

In the 1970s, in the early days of recombinant DNA research, a group of biologists led by Paul Berg formulated the 'Asilomar Declaration', advocating a moratorium on certain types of experiments. In retrospect, this move was perhaps over-cautious, but it seemed an encouraging precedent. However, it is surely far less likely that similar self-regulation could be achieved today. The research community is far larger, far more broadly international, and far more influenced by commercial pressures. One fears that, whatever regulatory regime is established on prudential or ethical grounds, anything that can be done will be done, somewhere (cf. the failure to enforce drug laws).

The physicist Freeman Dyson foresees a time when children will be able to design and create new organisms just as routinely as his generation played

with chemistry sets. I think this prospect is comfortably beyond the 'SF fringe', but were even part of this scenario to come about, our ecology (and even our species) surely would not long survive unscathed. (The consequences are so unpredictable and potentially widespread that it is unlikely that a bioterror event would be triggered by extremist groups with well-defined political aims. But such concerns would not give pause to an eco-fanatic who believed that 'Gaia' was being threatened by the presence of too many humans in the world. Most devastating would be a potentially fatal virus that was readily transmissible and had a long latency period).

Those of us with cushioned lives in the developed world fret too much about minor hazards: improbable air crashes, carcinogens in food, low radiation doses, and so forth. But we are less secure than we think. We (and our political masters) are in denial about catastrophic scenarios. These could be triggered as suddenly as the 2008 financial crisis; or they could develop insidiously. The worst have thankfully not yet happened – indeed they probably won't. But if an event is potentially catastrophic, it is worth paying a substantial premium to safeguard or insure against even if it is unlikely – just as we take out fire insurance on our house. Society could be dealt shattering blows by misapplication of technology that exists already, or that we can confidently expect within the next 20 years. It is, however, unrealistic to expect that we can ever be fully secure against bioerror and bioterror: risks would remain that cannot be eliminated except by measures that are themselves unpalatable, such as intrusive universal surveillance.

Looking forward to mid-century

But it's the trends in coming decades that should make us even more anxious. So I'll venture a word about these – but a tentative word, because scientists have a rotten record as forecasters. Lord Rutherford, the greatest nuclear physicist of his time, said in the 1930s that nuclear energy was 'moonshine'. One of my predecessors as Astronomer Royal said, as late as the 1950s, that space travel was 'utter bilge'. My own crystal ball is very cloudy.

We can predict confidently that in the latter part of the 21st century the world will be warmer and more crowded. But we can't predict how our lives might then have been changed by novel technologies. After all, the rapid societal transformation brought about by the smartphone, the Internet and their ancillaries would have seemed magic even 20 years ago. So, looking several decades ahead we must keep our minds open, or at least ajar, to prospects that may now seem in the realm of science fiction.

My own expertise is in astronomy and space technology. Colleagues may therefore think I worry specially about asteroid impacts. I don't. Indeed the

threat from asteroids is one of the few that we can quantify: we know roughly how many objects are on Earth-crossing orbits; and we know what the consequences would be from impacts of bodies with different sizes. Every few million years, there would be impact of a body a few kilometres across, causing global catastrophe – we have about one chance in 100000 that this will happen in our lifetime. But there is of course a higher chance of smaller impacts that would cause regional or local devastation. A body (say) 300 metres across, if it fell into the Atlantic, would produce huge tsunamis that would devastate the East Coast of the US, as well as much of Europe. An impact in Siberia in 1908 released energy equivalent to 5 megatons; a widely-reported impact last year was only a few times less powerful and such events happen, somewhere on Earth, about once a year.

Can we be forewarned of these impacts? The answer is already yes for the really big and rare ones – those triggered by bodies more than a kilometre across. However, only 1 percent of asteroids between 50 and 100 meters across have so far been detected. That is why I support the B612 project, spearheaded by former astronaut Ed Lu. The aim of this project is to put an infrared telescope in solar orbit to catalogue a million asteroids and monitor their orbits. With a few years' forewarning of where on Earth the impact would occur, action could be taken to mitigate its consequences on human populations by evacuating the most vulnerable areas. But what is even better news is that during this century we could develop the technology to protect us from impacts. A 'nudge', imparted a few years before the threatened impact, would only need to change an asteroid's velocity by a millimetre per second in order to deflect its path away from the Earth.

By 2100, groups of pioneers may have established bases independent from the Earth – on Mars, or maybe on asteroids. Whatever ethical constraints we impose here on the ground, we should surely wish these adventurers good luck in genetically modifying their progeny to adapt to alien environments. This might be the first step towards divergence into a new species: the beginning of the post-human era. And it would also ensure that advanced life would survive, even if the worst conceivable catastrophe befell our planet. But don't ever expect mass emigration from Earth. Nowhere in our Solar system offers an environment even as clement as the Antarctic or the top of Everest. Space doesn't offer an escape from Earth's problems.

The scope of biotechnology, and its consequent risks, will surely become more acute with each decade. But what about another fast-advancing technology: robotics and machine intelligence? Computers will surely vastly enhance our logical or mathematical skills, and perhaps even our creativity. We may be able to 'plug in' extra memory, or acquire language skills by di-

rect input into the brain (which would lead to a specially disquieting new form of inequality if such mental augmentations were available only to a privileged few). Even back in the 1990s IBM's 'Deep Blue' beat Kasparov, the world chess champion. More recently the Watson computer won a TV gameshow. Advances in software and sensors have been slower than in number-crunching capacity. Computers still can't match the facility of even a three-year-old child in telling a dog from a cat, or moving the pieces on a real chessboard. They can't tie your shoelaces or cut your toenails. But machine learning is advancing apace.

Once computers can observe and interpret their environment as adeptly as we do through our eyes and other sense organs, their far faster 'thoughts' and reactions could give them an advantage over us. [This will incidentally raise challenging ethical issues. We generally accept an obligation to ensure that other human beings (and indeed at least some animal species) can fulfil their 'natural' potential. Will we have the same duty to sophisticated robots, our own creations? Should we feel guilty about exploiting them, or if they are underemployed, frustrated, or bored?]

Moreover, we may need really to confront some science fiction scenarios – dumb autonomous robots 'going rogue', a 'supercomputer' with analytical powers offering its controller dominance of international finance, or a network that could develop a mind of its own and threaten us all. Be that as it may, by the end of this century, our society and its economy will be deeply changed by autonomous robots, but we should hope that this remain as 'idiot savants' rather than displaying full human capabilities. But can we be confident that machines will remain so limited? As early as the 1960s the British mathematician I.J. Good pointed out that a superintelligent robot (were it sufficiently versatile) could be the last invention that humans need ever make. Once machines have surpassed human capabilities, they could themselves design and assemble a new generation of even more intelligent ones, as well as an array of robotic fabricators that could transform the world physically.

Environmental threats and technological solutions

It is clear from other contributions to this conference that humanity's collective 'footprint' is threatening our finite planet's ecology. We should worry about the burgeoning environmental impact of a growing population needing food and energy – aggravated because, hopefully those in the developing world will close the consumption gap with the more fortunate among us. These pressures will be heightened because the world will also be warmer – though we can't forecast by how much, and how threatening

climate change will by then be. 'Ecological shocks' could irreversibly degrade our biosphere.

Doom-laden predictions of environmental catastrophe made in the 1970s proved wide off the mark. Unsurprisingly, such memories engender scepticism about the worst-case climatic and environmental projections. But the hazards may merely have been postponed. Climate change could plainly be devastating if the more pessimistic projections are borne out. Straightforward physics tells us that the anthropogenic rise in atmospheric CO2 (which is itself uncontroversial) will itself induce a long-term greenhouse warming: a rise of just over one degree Centigrade if CO2 doubles. This is superimposed on all the other complicated effects that make climate fluctuate. One degree may not seem much, but the direct 'greenhouse' effect of steadily-rising CO2 is thought to be amplified by consequent changes in water vapour and other greenhouse gases. These effects, and the consequences of changing cloud cover, aren't so well understood. The 5th IPCC report presents a spread of projections, depending on how much this 'carbon sensitivity factor' enhances the blanketing by CO2. The mean temperature rise is just an index for a warming that's very non-uniform, and which induces complex changes in weather patterns. And the worst consequences entail long time lags – it takes decades for the oceans to adjust to a new equilibrium, and centuries for ice-sheets to melt completely. The most compelling argument for prioritising mitigation, in my view, is the small risk of a runaway 'worst case' (as discussed, for instance, by Peter Wadhams) rather than the consequences of the median IPCC projections.

These 'sustainability' issues are familiar – so is the inaction in dealing with them and moving towards a lower-carbon economy. The inaction stems from the tendency in all democracies for the urgent to trump the long-term, and the parochial to trump the global.

It's uncertain how rapidly the climate will change and what 'insurance premium' we should be willing to pay to avoid the worst-case scenarios. My pessimistic guess is political efforts to decarbonise energy production will continue to be torpid rather than effective, and that the CO2 concentration in the atmosphere will rise at an accelerating rate throughout the next 20 years. But by then, we'll know with far more confidence – perhaps from advanced computer modelling, but also from how much global temperatures have actually risen by then – just how strongly the feedback from water vapour and clouds amplifies the effect of CO2 itself in creating a 'greenhouse effect'. If the effect is strong, and the world consequently seems on a rapidly-warming trajectory into dangerous territory, there may be a pressure for 'panic measures'. These would have to involve a 'plan B' – being

fatalistic about continuing dependence on fossil fuels, but combating its effects by some form of geoengineering.

The 'greenhouse warming' could be counteracted by (for instance) putting reflecting aerosols in the upper atmosphere, or even vast sunshades in space. The political problems of such geoengineering may be overwhelming. There could be unintended side- effects. Moreover, the warming would return with a vengeance of the countermeasures were ever discontinued; and other consequences of rising CO_2 (especially the deleterious effects of ocean acidification) would be unchecked.

An alternative strategy would involve direct extraction of carbon from the atmosphere. This approach would be politically more acceptable – we'd essentially just be undoing the unwitting geoengineering we've done by burning fossil fuels. But it currently seems less practicable.

It seems right at least to study geoengineering – to clarify which options make sense and perhaps damp down undue optimism about a technical 'quick fix' of our climate. However, it already seems clear that it would be feasible and affordable to throw enough material into the stratosphere to change the world's climate – indeed what is scary is that this capacity might be within the resources of a single nation, or even a single corporation or individual. Geoengineering would be an utter political nightmare: not all nations would want to adjust the thermostat the same way. Very elaborate climatic modelling would be needed in order to calculate the regional impacts of such an intervention. It would be prudent to sort out the complex governance issues raised by 'Solar Radiation Management' – and to do this well before urgent pressures for action might build up.

Are there genuinely 'existential' threats?

The events I've described could present serious, even catastrophic, setbacks to our civilization, but wouldn't wipe us all out. They're extreme, but strictly speaking not 'existential'. Are there conceivable events that could threaten the *entire* Earth, and snuff out all life? Promethean concerns of this kind were raised by scientists working on the atomic bomb project during the Second World War. Could we be absolutely sure that a nuclear explosion wouldn't ignite all the world's atmosphere or oceans? Before the first bomb test in New Mexico, Hans Bethe and two colleagues addressed this issue; they convinced themselves that there was a large safety factor. We now know for certain that a single nuclear weapons, devastating though it is, can't trigger a nuclear chain reaction that would utterly destroy the Earth or its atmosphere.

But what about even more extreme experiments? Physicists were (in my view quite rightly) pressured by the media to address the speculative

'existential risks' that could be triggered by powerful accelerators that generate unprecedented concentrations of energy. Could physicists unwittingly convert the entire Earth into particles called 'strangelets' – or, even worse, trigger a 'phase transition' that would rip apart the fabric of space itself? Fortunately, reassurance could be offered: it was pointed out that cosmic ray collisions of much higher energies occur frequently in the Galaxy, but haven't ripped space apart. And cosmic rays have penetrated white dwarf and neutron stars without triggering their conversion into 'strangelets'.

But physicists should surely be circumspect and precautionary about carrying out experiments that generate conditions with no precedent even in the cosmos – just as biologists should avoid release of potentially-devastating genetically-modified pathogens.

So how risk-averse should we be? Some would argue that odds of 10 million to one against a global disaster would be good enough, because that is below the chance that, within the next year, an asteroid large enough to cause global devastation will hit the Earth. (This is like arguing that the extra carcinogenic effect of artificial radiation is acceptable if it doesn't so much as double the risk from natural radiation). But to some, even this limit may not seem stringent enough. We may become resigned to a natural risk (like asteroids or natural pollutants) that we can't do much about, but that doesn't mean that we should acquiesce in an extra avoidable risk of the same magnitude. Designers of nuclear power-stations have to convince regulators that the probability of a meltdown is less than one in a million per year. Applying the same standards, if there were a threat to the entire Earth, the public might properly demand assurance that the probability is below one in a billion – even one in a trillion – before sanctioning such an experiment. We may offer these odds against the Sun not rising tomorrow, or against a fair die giving 100 sixes in a row; but a scientist might seem over-presumptuous to place such extreme confidence in any theories about what happens when atoms are smashed together with unprecedented energy. If a congressional committee asked: 'Are you really claiming that there's less than one chance in a billion chance that you're wrong?' I'd feel uncomfortable saying yes.

But on the other hand, if you ask: "Could such an experiment reveal a transformative discovery that – for instance – provided a new source of energy for the world?" I'd again offer high odds against it. The issue is then the relative probability of these two unlikely event – one hugely beneficial, the other catastrophic. Innovation is always risky, but if we don't take these risks we may forgo disproportionate benefits. Undiluted application of the 'precautionary principle' has a manifest downside. As Freeman Dyson argued in an eloquent essay, there is 'the hidden cost of saying no'.

And, by the way, the priority that we should assign to avoiding truly existential disasters, even when their probability seems infinitesimal, depends on an ethical question posed by the philosopher Derek Parfitt, which is this. Consider two scenarios: scenario A wipes out 90 percent of humanity; scenario B wipes out 100 percent. How much worse is B than A? Some would say 10 percent worse: the body count is 10 percent higher. But others would say B was *incomparably* worse, because human extinction forecloses the existence of billions, even trillions, of future people – and indeed an open-ended post-human future.

And especially if you accept the latter viewpoint, you'll agree that existential catastrophes – even if you'd bet a billion to one against them – deserve more attention than they're getting, in order that we can guard against them. That's why some of us in Cambridge – both natural and social scientists – plan to inaugurate a research programme to compile a more complete register of these 'existential' risks, and to assess how to enhance resilience against the more credible ones.

Moreover, we shouldn't be complacent that all such probabilities are so miniscule. We've no grounds for assuming that human-induced threats worse than those on our current risk register are improbable: they are newly emergent, so we have a limited timebase for exposure to them and can't be sanguine that we would survive them for long. And we have zero grounds for confidence that we can survive the worst that future technologies could bring in their wake. Some scenarios that have been envisaged may indeed be science fiction; but others may be disquietingly real.

Technology bring with it great hopes, but also great fears. We mustn't forget an important maxim: the unfamiliar is not the same as the improbable.

The role of scientists and their academies

More should be done to assess, and then minimize, the extreme risks I've addressed in this paper. But though we live under their shadow, we can nonetheless surely be technological optimists. There seems no scientific impediment to achieving (with very high probability) a sustainable world, where all enjoy a lifestyle better than those in the 'west' do today. But I'm a political pessimist. The intractable politics and sociology – the gap between potentialities and what actually happens – engenders pessimism. Politicians look to their own voters – and the next election. Stockholders expect a pay-off in the short run. We downplay what's happening even now in far-away countries. And we discount too heavily the problems we'll leave for new generations. Without a broader perspective, the public will never be adequately motivated to stem the risk of environmental degradation; to pri-

oritise clean energy, and sustainable agriculture; and to handle the challenge posed by ever more powerful technology.

Now that the impact of their researches can be so much greater, scientists surely have a still deeper responsibility to engage with governments and society. Politicians need the best 'in house' scientific advice. But, more than that, choices on how technology is applied require wide public debate, and such debate must be leveraged by 'scientific citizens' – engaging, from all political perspectives, with the media, and with a public attuned to the scope and limit of science. They can do this via campaigning groups, via blogging and journalism, or through political activity and thereby catalyse a debate that is better-informed. And there is a role for international academies. (Indeed I recall with admiration the efforts of the PAS, under the leadership of Profs Chagas, Weisskopf, Perutz and others in the 1980s, to urge on heads of state the importance of reducing nuclear arsenals).

We need to realise that we're all on this crowded world together. We are stewards of a precious 'pale blue dot' in a vast cosmos – a planet with a future measured in billions of years, whose fate depends on humanity's collective actions. We must urge greater priority for long-term global issues on the political agenda. And our institutions must prioritise projects that are long-term in a political perspective, even if a mere instant in the history of our planet. We need to broaden our sympathies in both space and time and perceive ourselves as part of a long heritage, and stewards for an immense future. We must be guided by the best science – both natural science and social science – but also by values that science itself can never provide.

I started by quoting H.G. Wells. I'll finish with a quote from another scientific sage, the biologist Peter Medawar:

"The bells that toll for mankind are ... like the bells of Alpine cattle. They are attached to our own necks, and it must be our fault if they do not make a tuneful and melodious sound".