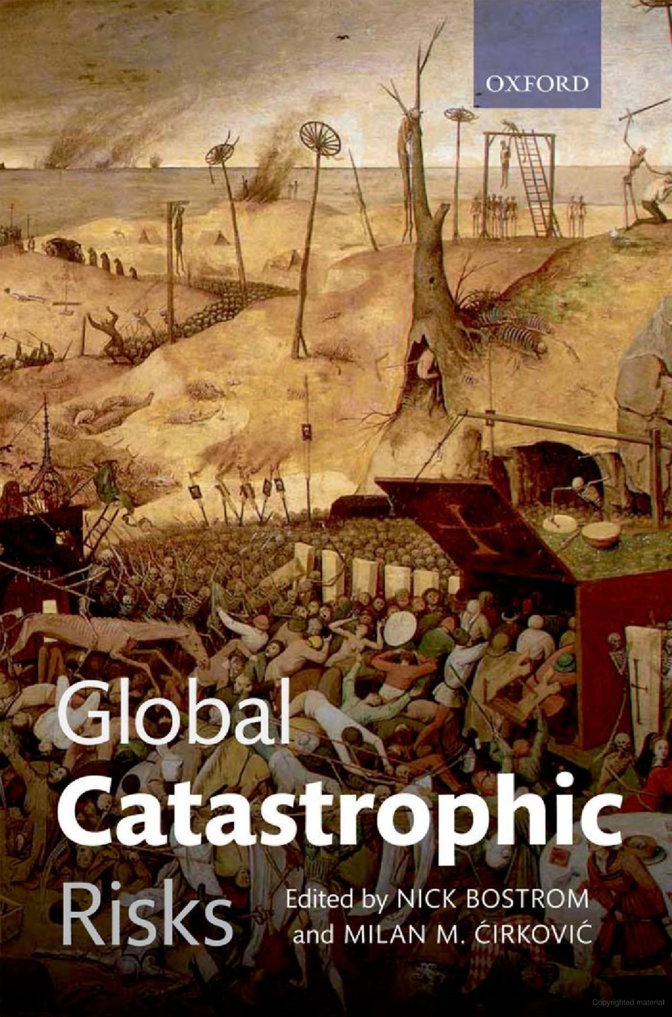

OXFORD

Global Catastrophic Risks

Edited by NICK BOSTROM
and MILAN M. ĆIRKOVIĆ

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2008

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire

ISBN 978–0–19–857050–9 (Hbk)

1 3 5 7 9 10 8 6 4 2

Contents

·1·

Introduction

Nick Bostrom and Milan M. Ćirković

1.1 Why?

The term 'global catastrophic risk' lacks a sharp definition. We use it to refer, loosely, to a risk that might have the potential to inflict serious damage to human well-being on a global scale. On this definition, an immensely diverse collection of events could constitute global catastrophes: potential candidates range from volcanic eruptions to pandemic infections, nuclear accidents to worldwide tyrannies, out-of-control scientific experiments to climatic changes, and cosmic hazards to economic collapse. With this in mind, one might well ask, what use is a book on global catastrophic risk? The risks under consideration seem to have little in common, so does 'global catastrophic risk' even make sense as a topic? Or is the book that you hold in your hands as ill-conceived and unfocused a project as a volume on 'Gardening, Matrix Algebra, and the History of Byzantium'?

We are confident that a comprehensive treatment of global catastrophic risk will be at least somewhat more useful and coherent than the above-mentioned imaginary title. We also believe that studying this topic is highly important. Although the risks are of various kinds, they are tied together by many links and commonalities. For example, for many types of destructive events, much of the damage results from second-order impacts on social order; thus the risks of social disruption and collapse are not unrelated to the risks of events such as nuclear terrorism or pandemic disease. Or to take another example, apparently dissimilar events such as large asteroid impacts, volcanic super-eruptions, and nuclear war would all eject massive amounts of soot and aerosols into the atmosphere, with significant effects on global climate. The existence of such causal linkages is one reason why it is can be sensible to study multiple risks together.

Another commonality is that many methodological, conceptual, and cultural issues crop up across the range of global catastrophic risks. If our interest lies in such issues, it is often illuminating to study how they play out in different contexts. Conversely, some general insights – for example, into the biases of human risk cognition – can be applied to many different risks and used to improve our assessments across the board.

Beyond these theoretical commonalities, there are also pragmatic reasons for addressing global catastrophic risks as a single field. Attention is scarce. Mitigation is costly. To decide how to allocate effort and resources, we must make comparative judgements. If we treat risks singly, and never as part of an overall threat profile, we may become unduly fixated on the one or two dangers that happen to have captured the public or expert imagination of the day, while neglecting other risks that are more severe or more amenable to mitigation. Alternatively, we may fail to see that some precautionary policy, while effective in reducing the particular risk we are focusing on, would at the same time create new hazards and result in an increase in the overall level of risk. A broader view allows us to gain perspective and can thereby help us to set wiser priorities.

The immediate aim of this book is to offer an introduction to the range of global catastrophic risks facing humanity now or expected in the future, suitable for an educated interdisciplinary readership. There are several constituencies for the knowledge presented. Academics specializing in one of these risk areas will benefit from learning about the other risks. Professionals in insurance, finance, and business – although usually preoccupied with more limited and imminent challenges – will benefit from a wider view. Policy analysts, activists, and laypeople concerned with promoting responsible policies likewise stand to gain from learning about the state of the art in global risk studies. Finally, anyone who is worried or simply curious about what could go wrong in the modern world might find many of the following chapters intriguing. We hope that this volume will serve as a useful introduction to all of these audiences. Each of the chapters ends with some pointers to the literature for those who wish to delve deeper into a particular set of issues.

This volume also has a wider goal: to stimulate increased research, awareness, and informed public discussion about big risks and mitigation strategies. The existence of an interdisciplinary community of experts and laypeople knowledgeable about global catastrophic risks will, we believe, improve the odds that good solutions will be found and implemented to the great challenges of the twenty-first century.

1.2 Taxonomy and organization

Let us look more closely at what would, and would not, count as a global catastrophic risk. Recall that the damage must be serious, and the scale global. Given this, a catastrophe that caused 10,000 fatalities or 10 billion dollars worth of economic damage (e.g., a major earthquake) would not qualify as a global catastrophe. A catastrophe that caused 10 million fatalities or 10 trillion dollars worth of economic loss (e.g., an influenza pandemic) would count as a global catastrophe, even if some region of the world escaped unscathed. As for disasters falling between these points, the definition is vague. The stipulation of a precise cut-off does not appear needful at this stage.

Global catastrophes have occurred many times in history, even if we only count disasters causing more than 10 million deaths. A very partial list of examples might include the An Shi Rebellion (756–763), the Taiping Rebellion (1851–1864), and the famine of the Great Leap Forward in China, the Black Death in Europe, the Spanish flu pandemic, the two world wars, the Nazi genocides, the famines in British India, Stalinist totalitarianism, the decimation of the native American population through smallpox and other diseases following the arrival of European colonizers, probably the Mongol conquests, perhaps Belgian Congo – innumerable others could be added to the list depending on how various misfortunes and chronic conditions are individuated and classified.

We can roughly characterize the severity of a risk by three variables: its *scope* (how many people – and other morally relevant beings – would be affected), its *intensity* (how badly these would be affected), and its *probability* (how likely the disaster is to occur, according to our best judgement, given currently available evidence). Using the first two of these variables, we can construct a qualitative diagram of different types of risk (Fig. 1.1). (The probability dimension could be displayed along a *z*-axis were this diagram three-dimensional.)

The scope of a risk can be *personal* (affecting only one person), *local, global* (affecting a large part of the human population), or *trans-generational* (affecting

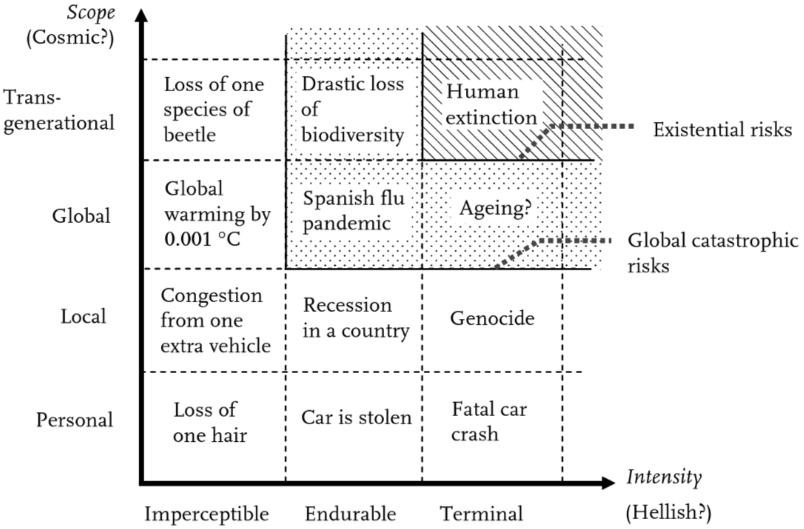

Fig. 1.1 Qualitative categories of risk. Global catastrophic risks are in the upper right part of the diagram. Existential risks form an especially severe subset of these.

not only the current world population but all generations that could come to exist in the future). The intensity of a risk can be classified as imperceptible (barely noticeable), endurable (causing significant harm but not destroying quality of life completely), or terminal (causing death or permanently and drastically reducing quality of life). In this taxonomy, global catastrophic risks occupy the four risks classes in the high-severity upper-right corner of the figure: a global catastrophic risk is of either global or trans-generational scope, and of either endurable or terminal intensity. In principle, as suggested in the figure, the axes can be extended to encompass conceptually possible risks that are even more extreme. In particular, trans-generational risks can contain a subclass of risks so destructive that their realization would not only affect or pre-empt future human generations, but would also destroy the potential of our future light cone of the universe to produce intelligent or self-aware beings (labelled 'Cosmic'). On the other hand, according to many theories of value, there can be states of being that are even worse than non-existence or death (e.g., permanent and extreme forms of slavery or mind control), so it could, in principle, be possible to extend the x-axis to the right as well (see Fig. 1.1 labelled 'Hellish').

A subset of global catastrophic risks is *existential risks*. An existential risk is one that threatens to cause the extinction of Earth-originating intelligent life or to reduce its quality of life (compared to what would otherwise have been possible) permanently and drastically.[1] Existential risks share a number of features that mark them out as deserving of special consideration. For example, since it is not possible to recover from existential risks, we cannot allow even one existential disaster to happen; there would be no opportunity to learn from experience. Our approach to managing such risks must be proactive. How much worse an existential catastrophe would be than a non-existential global catastrophe depends very sensitively on controversial issues in value theory, in particular how much weight to give to the lives of possible future persons.[2] Furthermore, assessing existential risks raises distinctive methodological problems having to do with observation selection effects and the need to avoid anthropic bias. One of the motives for producing this book is to stimulate more serious study of existential risks. Rather than limiting our focus to existential risk, however, we thought it better to lay a broader foundation of systematic thinking about big risks in general.

[1] (Bostrom, 2002, p. 381).

[2] For many aggregative consequentialist ethical theories, including but not limited to total utilitarianism, it can be shown that the injunction to *maximize expected value!* can be simplified – for all practical purposes – to the injunction to *minimize existential risk!* (Bostrom, 2003, p. 439). (Note, however, that aggregative consequentialism is threatened by the problem of infinitarian paralysis [Bostrom, 2007, p. 730].)

We asked our contributors to assess global catastrophic risks not only as they presently exist but also as they might develop over time. The temporal dimension is essential for a full understanding of the nature of the challenges we face. To think about how to tackle the risks from nuclear terrorism and nuclear war, for instance, we must consider not only the probability that something will go wrong within the next year, but also about how the risks will change in the future and the factors – such as the extent of proliferation of relevant technology and fissile materials – that will influence this. Climate change from greenhouse gas emissions poses no significant globally catastrophic risk now or in the immediate future (on the timescale of several decades); the concern is about what effects these accumulating emissions might have over the course of many decades or even centuries. It can also be important to anticipate hypothetical risks which will arise if and when certain possible technological developments take place. The chapters on nanotechnology and artificial intelligence are examples of such prospective risk analysis.

In some cases, it can be important to study scenarios which are almost certainly physically impossible. The hypothetical risk from particle collider experiments is a case in point. It is very likely that these experiments have no potential, whatever, for causing global disasters. The objective risk is probably zero, as believed by most experts. But just how confident can we be that there is no objective risk? If we are not certain that there is no objective risk, then there is a risk at least in a subjective sense. Such subjective risks can be worthy of serious consideration, and we include them in our definition of global catastrophic risks.

The distinction between objective and subjective (epistemic) risk is often hard to make out. The possibility of an asteroid colliding with Earth looks like a clear-cut example of objective risk. But suppose that in fact no sizeable asteroid is on collision course with our planet within a certain, sufficiently large interval of time. We might then say that there is no objective risk of an asteroid-caused catastrophe within that interval of time. Of course, we will not know that this is so until we have mapped out the trajectories of all potentially threatening asteroids and are able to calculate all perturbations, often chaotic, of those trajectories. In the meantime, we must recognize a risk from asteroids even though the risk might be purely subjective, merely reflecting our present state of ignorance. An empty cave can be similarly subjectively unsafe if you are unsure about whether a lion resides in it; and it can be rational for you to avoid the cave if you reasonably judge that the expected harm of entry outweighs the expected benefit.

In the case of the asteroid threat, we have access to plenty of data that can help us quantify the risk. We can estimate the probability of a catastrophic impact from statistics of past impacts (e.g., cratering data) and from observations sampling from the population of non-threatening asteroids. This particular risk, therefore, lends itself to rigorous scientific study, and the probability estimates we derive are fairly strongly constrained by hard evidence.[3]

For many other risks, we lack the data needed for rigorous statistical inference. We may also lack well-corroborated scientific models on which to base probability estimates. For example, there exists no rigorous scientific way of assigning a probability to the risk of a serious terrorist attack employing a biological warfare agent occurring within the next decade. Nor can we firmly establish that the risks of a global totalitarian regime arising before the end of the century are of a certain precise magnitude. It is inevitable that analyses of such risks will rely to a large extent on plausibility arguments, analogies, and subjective judgement.

Although more rigorous methods are to be preferred whenever they are available and applicable, it would be misplaced scientism to confine attention to those risks that are amenable to hard approaches.[4] Such a strategy would lead to many risks being ignored, including many of the largest risks confronting humanity. It would also create a false dichotomy between two types of risks – the 'scientific' ones and the 'speculative' ones – where, in reality, there is a continuum of analytic tractability.

We have, therefore, opted to cast our net widely. Although our topic selection shows some skew towards smaller risks that have been subject to more scientific study, we do have a range of chapters that tackle potentially large but more speculative risks. The page count allocated to a risk should not, of course, be interpreted as a measure of how seriously we believe the risk ought to be regarded. In some cases, we have seen it fit to have a chapter devoted to a risk that turns out to be quite small, because learning that a particular risk is small can be useful, and the procedures used to arrive at the conclusion might serve as a template for future risk research. It goes without saying that the exact composition of a volume like this is also influenced by many contingencies

---

[3] One can sometimes define something akin to objective physical probabilities ('chances') for deterministic systems, as is done, for example, in classical statistical mechanics, by assuming that the system is ergodic under a suitable course graining of its state space. But ergodicity is not necessary for there being strong scientific constraints on subjective probability assignments to uncertain events in deterministic systems. For example, if we have good statistics going back a long time showing that impacts occur on average once per thousand years, with no apparent trends or periodicity, then we have scientific reason – absent of more specific information – for assigning a probability of $\approx 0.1\%$ to an impact occurring within the next year, whether we think the underlying system dynamic is indeterministic, or chaotic, or something else.

[4] Of course, when allocating research effort it is legitimate to take into account not just how important a problem is but also the likelihood that a solution can be found through research. The drunk who searches for his lost keys where the light is best is not necessarily irrational; and a scientist who succeeds in something relatively unimportant may achieve more good than one who fails in something important.

beyond the editors' control and that perforce it must leave out more than it includes.[5]

We have divided the book into four sections:

Part I: Background
Part II: Risks from Nature
Part III: Risks from Unintended Consequences
Part IV: Risks from Hostile Acts

This subdivision into three categories of risks is for convenience only, and the allocation of a risk to one of these categories is often fairly arbitrary. Take earthquakes which might seem to be paradigmatically a 'Risk from Nature'. Certainly, an earthquake is a natural event. It would happen even if we were not around. Earthquakes are governed by the forces of plate tectonics over which human beings currently have no control. Nevertheless, the risk posed by an earthquake is, to a very large extent, a matter of human construction. Where we erect our buildings and how we choose to construct them strongly influence what happens when an earthquake of a given magnitude occurs. If we all lived in tents, or in earthquake-proof buildings, or if we placed our cities far from fault lines and sea shores, earthquakes would do little damage. On closer inspection, we thus find that the earthquake risk is very much a joint venture between Nature and Man. Or take a paradigmatically anthropogenic hazard such as nuclear weapons. Again we soon discover that the risk is not as disconnected from uncontrollable forces of nature as might at first appear to be the case. If a nuclear bomb goes off, how much damage it causes will be significantly influenced by the weather. Wind, temperature, and precipitation will affect the fallout pattern and the likelihood that a fire storm will break out: factors that make a big difference to the number of fatalities generated by the blast. In addition, depending on how a risk is defined, it may also over time transition from one category to another. For instance, the risk of starvation might once have been primarily a Risk from Nature, when the main causal factors were draughts or fluctuations in local prey population; yet in the contemporary world, famines tend to be the consequences of market failures, wars, and social breakdowns, whence the risk is now at least as much one of Unintended Consequences or of Hostile Acts.

1.3 Part I: Background

The objective of this part of the book is to provide general context and methodological guidance for thinking systematically and critically about global catastrophic risks.

---

[5] For example, the risk of large-scale conventional war is only covered in passing, yet would surely deserve its own chapter in a more ideally balanced page allocation.

We begin at the end, as it were, with Chapter 2 by Fred Adams discussing the long-term fate of our planet, our galaxy, and the Universe in general. In about 3.5 billion years, the growing luminosity of the sun will essentially have sterilized the Earth's biosphere, but the end of *complex* life on Earth is scheduled to come sooner, maybe 0.9–1.5 billon years from now. This is the default fate for life on our planet. One may hope that if humanity and complex technological civilization survives, it will long before then have learned to colonize space.

If some cataclysmic event were to destroy *Homo sapiens* and other higher organisms on Earth tomorrow, there does appear to be a window of opportunity of approximately one billion years for another intelligent species to evolve and take over where we left off. For comparison, it took approximately 1.2 billion years from the rise of sexual reproduction and simple multicellular organisms for the biosphere to evolve into its current state, and only a few million years for our species to evolve from its anthropoid ancestors. Of course, there is no guarantee that a rerun of evolution would produce anything like a human or a self-aware successor species.

If intelligent life does spread into space by harnessing the powers of technology, its lifespan could become extremely long. Yet eventually, the universe will wind down. The last stars will stop shining 100 trillion years from now. Later, matter itself will disintegrate into its basic constituents. By 10^{100} years from now even the largest black holes would have evaporated. Our present understanding of what will happen at this time scale and beyond is quite limited. The current best guess – but it is really no more than that – is that it is not just technologically difficult but physically impossible for intelligent information processing to continue beyond some finite time into the future. If so, extinction is not a question of whether, but when.

After this peek into the extremely remote future, it is instructive to turn around and take a brief peek at the distant past. Some past cataclysmic events have left traces in the geological record. There have been about fifteen mass extinctions in the last 500 million years, and five of these eliminated more half of all species then inhabiting the Earth. Of particular note is the Permian – Triassic extinction event, which took place some 251.4 million years ago. This 'mother of all mass extinctions' eliminated more than 90% of all species and many entire phylogenetic families. It took upwards of 5 million years for biodiversity to recover.

Impacts from asteroids and comets, as well as massive volcano eruptions, have been implicated in many of the mass extinctions of the past. Other causes, such as variations in the intensity of solar illumination, may in some cases have exacerbated stresses. It appears that all mass extinctions have been mediated by atmospheric effects such as changes the atmosphere's composition or temperature. It is possible, however, that we owe our existence to mass extinctions. In particular, the comet that hit Earth 65 million years ago, which is believed to have been responsible for the demise of the dinosaurs, might have been a sine qua non for the subsequent rise of *Homo sapiens* by clearing an ecological niche that could be occupied by large mammals, including our ancestors.

At least 99.9% of all species that have ever walked, crawled, flown, swum, or otherwise abided on Earth are extinct. Not all of these were eliminated in cataclysmic mass extinction events. Many succumbed in less spectacular doomsdays such as from competition by other species for the same ecological niche. Chapter 3 reviews the mechanisms of evolutionary change. Not so long ago, our own species co-existed with at least one other hominid species, the Neanderthals. It is believed that the lineages of *H. sapiens* and *H. neanderthalensis* diverged about 800,000 years ago. The Neanderthals manufactured and used composite tools such as handaxes. They did not reach extinction in Europe until 33,000 to 24,000 years ago, quite likely as a direct result of competition with *Homo sapiens*. Recently, the remains of what might have been another hominoid species, *Homo floresiensis* – nicknamed 'the hobbit' for its short stature – were discovered on an Indonesian island. *H. floresiensis* is believed to have survived until as recently as 12,000 years ago, although uncertainty remains about the interpretation of the finds. An important lesson of this chapter is that extinction of intelligent species *has* already happened on Earth, suggesting that it would be naïve to think it may not happen again.

From a naturalistic perspective, there is thus nothing abnormal about global cataclysms including species extinctions, although the characteristic time scales are typically large by human standards. James Hughes in Chapter 4 makes clear, however, the idea of cataclysmic endings often causes a peculiar set of cognitive tendencies to come into play, what he calls 'the millennial, utopian, or apocalyptic psychocultural bundle, a characteristic dynamic of eschatological beliefs and behaviours'. The millennial impulse is pancultural. Hughes shows how it can be found in many guises and with many common tropes from Europe to India to China, across the last several thousand years. 'We may aspire to a purely rational, technocratic analysis', Hughes writes, 'calmly balancing the likelihoods of futures without disease, hunger, work or death, on the one hand, against the likelihoods of worlds destroyed by war, plagues or asteroids, but few will be immune to millennial biases, positive or negative, fatalist or messianic'. Although these eschatological tropes can serve legitimate social needs and help to mobilize needed action, they easily become dysfunctional and contribute to social disengagement. Hughes argues that we need historically informed and vigilant self-interrogation to help us keep our focus on constructive efforts to address real challenges.

Even for an honest, truth-seeking, and well-intentioned investigator it is difficult to think and act rationally in regard to global catastrophic risks and existential risks. These are topics on which it seems especially difficult to remain sensible. In Chapter 5, Eliezer Yudkowsky observes as follows:

> Substantially larger numbers, such as 500 million deaths, and *especially* qualitatively different scenarios such as the extinction of the entire human species, seem to trigger a *different mode of thinking* – enter into a 'separate magisterium'. People who would never dream of hurting a child hear of an existential risk, and say, 'Well, maybe the human species doesn't really deserve to survive'.

Fortunately, if we are ready to contend with our biases, we are not left entirely to our own devices. Over the last few decades, psychologists and economists have developed an extensive empirical literature on many of the common heuristics and biases that can be found in human cognition. Yudkowsky surveys this literature and applies its frequently disturbing findings to the domain of large-scale risks that is the subject matter of this book. His survey reviews the following effects: availability; hindsight bias; black swans; the conjunction fallacy; confirmation bias; anchoring, adjustment, and contamination; the affect heuristic; scope neglect; calibration and overconfidence; and bystander apathy. It behooves any sophisticated contributor in the area of global catastrophic risks and existential risks – whether scientist or policy advisor – to be familiar with each of these effects and we all ought to give some consideration to how they might be distorting our judgements.

Another kind of reasoning trap to be avoided is anthropic bias. Anthropic bias differs from the general cognitive biases reviewed by Yudkowsky; it is more theoretical in nature and it applies more narrowly to only certain specific kinds of inference. Anthropic bias arises when we overlook relevant observation selection effects. An observation selection effect occurs when our evidence has been 'filtered' by the precondition that a suitably positioned observer exists to have the evidence, in such a way that our observations are unrepresentatively sampled from the target domain. Failure to take observation effects into account correctly can result in serious errors in our probabilistic evaluation of some of the relevant hypotheses. Milan Ćirković, in Chapter 6, reviews some applications of observation selection theory that bear on global catastrophic risk and particularly existential risk. Some of these applications are fairly straightforward albeit not always obvious. For example, the tempting inference that certain classes of existential disaster must be highly improbable because they have never occurred in the history of our species or even in the history of life on Earth must be resisted. We are bound to find ourselves in one of those places and belonging to one of those intelligent species which have not yet been destroyed, whether planet or species-destroying disasters are common or rare: for the alternative possibility – that *our* planet has been destroyed or *our* species extinguished – is something that is unobservable for us, per definition. Other applications of anthropic reasoning – such as the Carter–Leslie Doomsday argument – are of disputed validity, especially in their generalized forms, but nevertheless worth knowing about. In some applications, such as the simulation argument, surprising constraints are revealed on what we can coherently assume about humanity's future and our place in the world.

There are professional communities that deal with risk assessment on a daily basis. The subsequent two chapters present perspectives from the systems engineering discipline and the insurance industry, respectively.

In Chapter 7, Yacov Haimes outlines some flexible strategies for organizing our thinking about risk variables in complex systems engineering projects. What knowledge is needed to make good risk management decisions? Answering this question, Haimes says, 'mandates seeking the "truth" about the unknowable complex nature of emergent systems; it requires intellectually bias-free modellers and thinkers who are empowered to experiment with a multitude of modelling and simulation approaches and to collaborate for appropriate solutions'. Haimes argues that organizing the analysis around the measure of the expected value of risk can be too constraining. Decision makers often prefer a more fine-grained decomposition of risk that allows them to consider separately the probability of outcomes in different severity ranges, using what Haimes calls 'the partitioned multi-objective risk method'.

Chapter 8, by Peter Taylor, explores the connections between the insurance industry and global catastrophic risk. Insurance companies help individuals and organizations mitigate the financial consequences of risk, essentially by allowing risks to be traded and shared. Peter Taylor argues that the extent to which global catastrophic risks can be privately insured is severely limited for reasons having to do with both their scope and their type.

Although insurance and reinsurance companies have paid relatively scant attention to global catastrophic risks, they have accumulated plenty of experience with smaller risks. Some of the concepts and methods used can be applied to risks at any scale. Taylor highlights the importance of the concept of *uncertainty*. A particular stochastic model of phenomena in some domain (such as earthquakes) may entail a definite probability distribution over possible outcomes. However, in addition to the chanciness described by the model, we must recognize two further sources of uncertainty. There is usually uncertainty in the values of the parameters that we feed into the model. On top of that, there is uncertainty about whether the model we use does, in fact, correctly describe the phenomena in the target domain. These higher-level uncertainties are often impossible to analyse in a statistically rigorous way. Analysts who strive for objectivity and who are expected to avoid making 'un-scientific' assumptions that they cannot justify face a temptation to ignore these subjective uncertainties. But such scientism can lead to disastrous misjudgements. Taylor argues that the distortion is often greatest at the tail end of exceedance probability curves, leading to an underestimation of the risk of extreme events.

Taylor also reports on two recent survey studies of perceived risk. One of these, conducted by Swiss Re in 2005, asked executives of multinationals about which risks to their businesses' financials were of greatest concern to them. Computer-related risk was rated as the highest priority risk, followed by foreign trade, corporate governance, operational/facility, and liability risk. Natural disasters came in seventh place, and terrorism in tenth place. It appears that, as far as financial threats to individual corporations are concerned, global catastrophic risks take the backseat to more direct and narrowly focused business hazards. A similar exercise, but with broader scope, is carried out annually by the World Economic Forum. Its 2007 Global Risk report classified risks by likelihood and severity based on opinions solicited from business leaders, economists, and academics. Risks were evaluated with a 10-year time frame. Two risks were given a severity rating of 'more than 1 trillion USD', namely, asset price collapse (10–20%) and retrenchment from globalization (1–5%). When severity was measured in number of deaths rather than economic losses, the top three risks were pandemics, developing world disease, and interstate and civil war. (Unfortunately, several of the risks in this survey were poorly defined, making it hard to interpret the reported opinions – one moral here being that, if one wishes to assign probabilities to risks or rank them according to severity or likelihood, an essential first step is to present clear definitions of the risks that are to be evaluated.[6])

The Background part of the book ends with a discussion by Richard Posner on some challenges for public policy in Chapter 9. Posner notes that governmental action to reduce global catastrophic risk is often impeded by the short decision horizons of politicians with their limited terms of office and the many competing demands on their attention. Furthermore, mitigation of global catastrophic risks is often costly and can create a free-rider problem. Smaller and poorer nations may drag their heels in the hope of taking a free ride on larger and richer countries. The more resourceful countries, in turn, may hold back because of reluctance to reward the free riders.

Posner also looks at several specific cases, including tsunamis, asteroid impacts, bioterrorism, accelerator experiments, and global warming, and considers some of the implications for public policy posed by these risks. Although rigorous cost–benefit analyses are not always possible, it is nevertheless important to attempt to quantify probabilities, potential harms, and the costs of different possible countermeasures, in order to determine priorities and optimal strategies for mitigation. Posner suggests that when

---

[6] For example, the risk 'Chronic disease in the developed world' is defined as 'Obesity, diabetes and cardiovascular diseases become widespread; healthcare costs increase; resistant bacterial infections rise, sparking class-action suits and avoidance of hospitals'. By most standards, obesity, diabetes, and cardiovascular disease are *already* widespread. And by *how much* would healthcare costs have to increase to satisfy the criterion? It may be impossible to judge whether this definition was met even after the fact and with the benefit of hindsight.

a precise probability of some risk cannot be determined, it can sometimes be informative to consider – as a rough heuristic – the ‘implied probability’ suggested by current expenditures on mitigation efforts compared to the magnitude of harms that would result if a disaster materialized. For example, if we spend one million dollars per year to mitigate a risk which would create 1 billion dollars of damage, we may estimate that current policies implicitly assume that the annual risk of the disaster is of the order of 1/1000. If this implied probability seems too small, it might be a sign that we are not spending enough on mitigation.[7] Posner maintains that the world is, indeed, under-investing in mitigation of several global catastrophic risks.

1.4 Part II: Risks from nature

Volcanic eruptions in recent historical times have had measurable effects on global climate, causing global cooling by a few tenths of one degree, the effect lasting perhaps a year. But as Michael Rampino explains in Chapter 10, these eruptions pale in comparison to the largest recorded eruptions. Approximately 75,000 years ago, a volcano erupted in Toba, Indonesia, spewing vast volumes of fine ash and aerosols into the atmosphere, with effects comparable to nuclear-winter scenarios. Land temperatures globally dropped by 5–15°C, and ocean-surface cooling of ≈2–6°C might have extended over several years. The persistence of significant soot in the atmosphere for one to three years might have led to a cooling of the climate lasting for decades (because of climate feedbacks such as increased snow cover and sea ice causing more of the sun’s radiation to be reflected back into space). The human population appears to have gone through a bottleneck at this time, according to some estimates dropping as low as approximately five hundred reproducing females in a world population of approximately 4000 individuals. On the Toba catastrophe theory, the population decline was caused by the super-eruption, and the human species was teetering on the brink of extinction. This is perhaps the worst disaster that has ever befallen the human species, at least if severity is measured by how close to terminal was the outcome.

More than twenty super-eruption sites for the last two million years have been identified. This would suggest that, on average, a super-eruption occurs at least once every 50,000 years. However, there may well have been additional super-eruptions that have not yet been identified in the geological record.

---

[7] This heuristic is only meant to be a first stab at the problem. It is obviously not generally valid. For example, if one million dollars is sufficient to take all the possible precautions, there is no reason to spend more on the risk even if we think that its probability is much greater than 1/1000. A more careful analysis would consider the marginal returns on investment in risk reduction.

The global damage from super-volcanism would come chiefly from its climatic effects. The volcanic winter that would follow such an eruption would cause a drop in agricultural productivity which could lead to mass starvation and consequent social upheavals. Rampino's analysis of the impacts of super-volcanism is also relevant to the risks of nuclear war and asteroid or meteor impacts. Each of these would involve soot and aerosols being injected into the atmosphere, cooling the Earth's climate.

Although we have no way of preventing a super-eruption, there are precautions that we could take to mitigate its impacts. At present, a global stockpile equivalent to a two-month supply of grain exists. In a super-volcanic catastrophe, growing seasons might be curtailed for several years. A larger stockpile of grain and other foodstuffs, while expensive to maintain, would provide a buffer for a range of catastrophe scenarios involving temporary reductions in world agricultural productivity.

The hazard from comets and meteors is perhaps the best understood of all global catastrophic risks (which is not to deny that significant uncertainties remain). Chapter 11, by William Napier, explains some of the science behind the impact hazards: where comets and asteroids come from, how frequently impacts occur, and what the effects of an impact would be. To produce a civilization-disrupting event, an impactor would need a diameter of at least one or two kilometre. A ten kilometre impactor would, it appears, have a good chance of causing the extinction of the human species. But even sub-kilometre impactors could produce damage reaching the level of global catastrophe, depending on their composition, velocity, angle, and impact site.

Napier estimates that 'the per capita impact hazard is at the level associated with the hazards of air travel and the like'. However, funding for mitigation is meager compared to funding for air safety. The main effort currently underway to address the impact hazard is the *Spaceguard* project, which receives about four million dollars per annum from NASA besides in-kind and voluntary contributions from others. *Spaceguard* aims to find 90% of near-Earth asteroids larger than one kilometre by the end of 2008. Asteroids constitute the largest portion of the threat from near-Earth objects (and are easier to detect than comets) so when the project is completed, the subjective probability of a large impact will have been reduced considerably – unless, of course, it were discovered that some asteroid has a date with our planet in the near future, in which case the probability would soar.

Some preliminary study has been done of how a potential impactor could be deflected. Given sufficient advance warning, it appears that the space technology needed to divert an asteroid could be developed. The cost of producing an effective asteroid defence would be much greater than the cost of searching for potential impactors. However, if a civilization-destroying wrecking ball were found to be swinging towards the Earth, virtually any expense would be justified to avert it before it struck.

Asteroids and comets are not the only potential global catastrophic threats from space. Other cosmic hazards include global climatic change from fluctuations in solar activity, and very large fluxes from radiation and cosmic rays from supernova explosions or gamma ray bursts. These risks are examined in Chapter 12 by Arnon Dar. The findings on these risks are favourable: the risks appear to be very small. No particular response seems indicated at the present time beyond continuation of basic research.[8]

1.5 Part III: Risks from unintended consequences

We have already encountered climate change – in the form of sudden global cooling – as a destructive modality of super-eruptions and large impacts (as well as a possible consequence of large-scale nuclear war, to be discussed later). Yet it is the risk of gradual global warming brought about by greenhouse gas emissions that has most strongly captured the public imagination in recent years. Anthropogenic climate change has become the poster child of global threats. Global warming commandeers a disproportionate fraction of the attention given to global risks.

Carbon dioxide and other greenhouse gases are accumulating in the atmosphere, where they are expected to cause a warming of Earth's climate and a concomitant rise in seawater levels. The most recent report by the United Nations' Intergovernmental Panel on Climate Change (IPCC), which represents the most authoritative assessment of current scientific opinion, attempts to estimate the increase in global mean temperature that would be expected by the end of this century under the assumption that no efforts at mitigation are made. The final estimate is fraught with uncertainty because of uncertainty about what the default rate of emissions of greenhouse gases will be over the century, uncertainty about the climate sensitivity parameter, and uncertainty about other factors. The IPCC, therefore, expresses its assessment in terms of six different climate scenarios based on different models and different assumptions. The 'low' model predicts a mean global warming of +1.8°C (uncertainty range 1.1–2.9°C); the 'high' model predicts warming by +4.0°C (2.4–6.4°C). Estimated sea level rise predicted by the two most extreme scenarios of the six considered is 18–38 cm, and 26–59 cm, respectively.

Chapter 13, by David Frame and Myles Allen, summarizes some of the basic science behind climate modelling, with particular attention to the low-probability high-impact scenarios that are most relevant to the focus of this book. It is, arguably, this range of extreme scenarios that gives the greatest

---

[8] A comprehensive review of space hazards would also consider scenarios involving contact with intelligent extraterrestrial species or contamination from hypothetical extraterrestrial microorganisms; however, these risks are outside the scope of Chapter 12.

cause for concern. Although their likelihood seems very low, considerable uncertainty still pervades our understanding of various possible feedbacks that might be triggered by the expected climate forcing (recalling Peter Taylor's point, referred to earlier, about the importance of taking parameter and model uncertainty into account). David Frame and Myles Allen also discuss mitigation policy, highlighting the difficulties of setting appropriate mitigation goals given the uncertainties about what levels of cumulative emissions would constitute 'dangerous anthropogenic interference' in the climate system.

Edwin Kilbourne reviews some historically important pandemics in Chapter 14, including the distinctive characteristics of their associated pathogens, and discusses the factors that will determine the extent and consequences of future outbreaks.

Infectious disease has exacted an enormous toll of suffering and death on the human species throughout history and continues to do so today. Deaths from infectious disease currently account for approximately 25% of all deaths worldwide. This amounts to approximately 15 million deaths per year. About 75% of these deaths occur in Southeast Asia and sub-Saharan Africa. The top five causes of death due to infectious disease are upper respiratory infection (3.9 million deaths), HIV/AIDS (2.9 million), diarrhoeal disease (1.8 million), tuberculosis (1.7 million), and malaria (1.3 million).

Pandemic disease is indisputably one of the biggest global catastrophic risks facing the world today, but it is not always accorded its due recognition. For example, in most people's mental representation of the world, the influenza pandemic of 1918–1919 is almost completely overshadowed by the concomitant World War I. Yet although the WWI is estimated to have directly caused about 10 million military and 9 million civilian fatalities, the Spanish flu is believed to have killed at least 20–50 million people. The relatively low 'dread factor' associated with this pandemic might be partly due to the fact that only approximately 2–3% of those who got sick died from the disease. (The total death count is vast because a large percentage of the world population was infected.)

In addition to fighting the major infectious diseases currently plaguing the world, it is vital to remain alert to emerging new diseases with pandemic potential, such as SARS, bird flu, and drug-resistant tuberculosis. As the World Health Organization and its network of collaborating laboratories and local governments have demonstrated repeatedly, decisive early action can sometimes nip an emerging pandemic in the bud, possibly saving the lives of millions.

We have chosen to label pandemics a 'risk from unintended consequences' even though most infectious diseases (exempting the potential of genetically engineered bioweapons) in some sense arise from nature. Our rationale is that the evolution as well as the spread of pathogens is highly dependent on human civilization. The worldwide spread of germs became possible only after all the inhabited continents were connected by travel routes. By now, globalization in the form of travel and trade has reached such an extent that a highly contagious disease could spread to virtually all parts of the world within a matter of days or weeks. Kilbourne also draws attention to another aspect of globalization as a factor increasing pandemic risk: homogenization of peoples, practices, and cultures. The more the human population comes to resemble a single homogeneous niche, the greater the potential for a single pathogen to saturate it quickly. Kilbourne mentions the 'one rotten apple syndrome', resulting from the mass production of food and behavioural fads:

> If one contaminated item, apple, egg or, most recently, spinach leaf carries a billion bacteria – not an unreasonable estimate – and it enters a pool of cake mix constituents then packaged and sent to millions of customers nationwide, a bewildering epidemic may ensue.

Conversely, cultural as well as genetic diversity reduces the likelihood that any single pattern will be adopted universally before it is discovered to be dangerous – whether the pattern be virus RNA, a dangerous new chemical or material, or a stifling ideology.

By contrast to pandemics, artificial intelligence (AI) is not an ongoing or imminent global catastrophic risk. Nor is it as uncontroversially a serious cause for concern. However, from a long-term perspective, the development of general artificial intelligence exceeding that of the human brain can be seen as one of the main challenges to the future of humanity (arguably, even as *the* main challenge). At the same time, the successful deployment of friendly superintelligence could obviate many of the other risks facing humanity. The title of Chapter 15, 'Artificial Intelligence as a positive and negative factor in global risk', reflects this ambivalent potential.

As Eliezer Yudkowsky notes, the prospect of superintelligent machines is a difficult topic to analyse and discuss. Appropriately, therefore, he devotes a substantial part of his chapter to clearing common misconceptions and barriers to understanding. Having done so, he proceeds to give an argument for giving serious consideration to the possibility that radical superintelligence could erupt very suddenly – a scenario that is sometimes referred to as the 'Singularity hypothesis'. Claims about the steepness of the transition must be distinguished from claims about the timing of its onset. One could believe, for example, that it will be a long time before computers are able to match the general reasoning abilities of an average human being, but that once that happens, it will only take a short time for computers to attain radically superhuman levels.

Yudkowsky proposes that we conceive of a superintelligence as an enormously powerful optimization process: 'a system which hits small targets in large search spaces to produce coherent real-world effects'. The superintelligence will be able to manipulate the world (including human beings) in such a way as to achieve its goals, whatever those goals might be.

To avert disaster, it would be necessary to ensure that the superintelligence is endowed with a 'Friendly' goal system: that is, one that aligns the system's goals with genuine human values.

Given this set-up, Yudkowsky identifies two different ways in which we could fail to build Friendliness into our AI: philosophical failure and technical failure. The warning against philosophical failure is basically that we should be careful what we wish for because we might get it. We might designate a target for the AI which at first sight seems like a nice outcome but which in fact is radically misguided or morally worthless. The warning against technical failure is that we might fail to get what we wish for, because of faulty implementation of the goal system or unintended consequences of the way the target representation was specified. Yudkowsky regards both of these possible failure modes as very serious existential risks and concludes that it is imperative that we figure out how to build Friendliness into a superintelligence before we figure out how to build a superintelligence.

Chapter 16 discusses the possibility that the experiments that physicists carry out in particle accelerators might pose an existential risk. Concerns about such risks prompted the director of the Brookhaven Relativistic Heavy Ion Collider to commission an official report in 2000. Concerns have since resurfaced with the construction of more powerful accelerators such as CERN's Large Hadron Collider. Following the Brookhaven report, Frank Wilczek distinguishes three catastrophe scenarios:

1. Formation of tiny black holes that could start accreting surrounding matter, eventually swallowing up the entire planet.
2. Formation of negatively charged stable strangelets which could catalyse the conversion of all the ordinary matter on our planet into strange matter.
3. Initiation of a phase transition of the vacuum state, which would propagate outward in all directions at near light speed and destroy not only our planet but the entire accessible part of the universe.

Wilczek argues that these scenarios are exceedingly unlikely on various theoretical grounds. In addition, there is a more general argument that these scenarios are extremely improbable which depends less on arcane theory. Cosmic rays often have energies far greater than those that will be attained in any of the planned accelerators. Such rays have been bombarding the Earth's atmosphere (and the moon and other astronomical objects) for billions of years without a single catastrophic effect having been observed. Assuming that collisions in particle accelerators do not differ in any unknown relevant respect from those that occur in the wild, we can be very confident in the safety of our accelerators.

By everyone's reckoning, it is highly improbable that particle accelerator experiments will cause an existential disaster. The question is *how* improbable? And what would constitute an 'acceptable' probability of an existential disaster?

In assessing the probability, we must consider not only how unlikely the outcome seems given our best current models but also the possibility that our best models and calculations might be flawed in some as-yet unrealized way. In doing so we must guard against overconfidence bias (compare Chapter 5 on biases). Unless we ourselves are technically expert, we must also take into account the possibility that the experts on whose judgements we rely might be consciously or unconsciously biased.[9] For example, the physicists who possess the expertise needed to assess the risks from particle physics experiments are part of a professional community that has a direct stake in the experiments going forward. A layperson might worry that the incentives faced by the experts could lead them to err on the side of downplaying the risks.[10] Alternatively, some experts might be tempted by the media attention they could get by playing up the risks. The issue of how much and in which circumstances to trust risk estimates by experts is an important one, and it arises quite generally with regard to many of the risks covered in this book.

Chapter 17 (by Robin Hanson) from Part III on Risks from unintended consequences focuses on social collapse as a devastation multiplier of other catastrophes. Hanson writes as follows:

> The main reason to be careful when you walk up a flight of stairs is not that you might slip and have to retrace one step, but rather that the first slip might cause a second slip, and so on until you fall dozens of steps and break your neck. Similarly we are concerned about the sorts of catastrophes explored in this book not only because of their terrible direct effects, but also because they may induce an even more damaging collapse of our economic and social systems.

This argument does not apply to some of the risks discussed so far, such as those from particle accelerators or the risks from superintelligence as envisaged by Yudkowsky. In those cases, we may be either completely safe or altogether doomed, with little probability of intermediary outcomes. But for many other types of risk – such as windstorms, tornados, earthquakes, floods, forest fires, terrorist attacks, plagues, and wars – a wide range of outcomes are possible, and the potential for social disruption or even social collapse constitutes a major part of the overall hazard. Hanson notes that many of these risks appear to follow a power law distribution. Depending on the characteristic exponent of such a power law distribution, most of the damage expected from a given

[9] Even if we ourselves are expert, we must still be alert to unconscious biases that may influence our judgment (e.g., anthropic biases, see Chapter 6).

[10] If experts anticipate that the public will not quite trust their reassurances, they might be led to try to sound even more reassuring than they would have if they had believed that the public would accept their claims at face value. The public, in turn, might respond by discounting the experts' verdicts even more, leading the experts to be even more wary of fuelling alarmist overreactions. In the end, experts might be reluctant to acknowledge any risk at all for fear of a triggering a hysterical public overreaction. Effective risk communication is a tricky business, and the trust that it requires can be hard to gain and easy to lose.

type of risk may consist either of frequent small disturbances or of rare large catastrophes. Car accidents, for example, have a large exponent, reflecting the fact that most traffic deaths occur in numerous small accidents involving one or two vehicles. Wars and plagues, by contrast, appear to have small exponents, meaning that most of the expected damage occurs in very rare but very large conflicts and pandemics.

After giving a thumbnail sketch of economic growth theory, Hanson considers an extreme opposite of economic growth: sudden reduction in productivity brought about by escalating destruction of social capital and coordination. For example, 'a judge who would not normally consider taking a bribe may do so when his life is at stake, allowing others to expect to get away with theft more easily, which leads still others to avoid making investments that might be stolen, and so on. Also, people may be reluctant to trust bank accounts or even paper money, preventing those institutions from functioning.' The productivity of the world economy depends both on scale and on many different forms of capital which must be delicately coordinated. We should be concerned that a relatively small disturbance (or combination of disturbances) to some vulnerable part of this system could cause a far-reaching unraveling of the institutions and expectations upon which the global economy depends.

Hanson also offers a suggestion for how we might convert some existential risks into non-existential risks. He proposes that we consider the construction of one or more continuously inhabited refuges – located, perhaps, in a deep mineshaft, and well-stocked with supplies – which could preserve a small but sufficient group of people to repopulate a post-apocalyptic world. It would obviously be preferable to prevent altogether catastrophes of a severity that would make humanity's survival dependent on such modern-day 'Noah's arks'; nevertheless, it might be worth exploring whether some variation of this proposal might be a cost-effective way of somewhat decreasing the probability of human extinction from a range of potential causes.[11]

1.6 Part IV: Risks from hostile acts

The spectre of nuclear Armageddon, which so haunted the public imagination during the Cold War era, has apparently entered semi-retirement. The number of nuclear weapons in the world has been reduced to half, from a Cold War high of 65,000 in 1986 to approximately 26,000 in 2007, with approximately

---

[11] Somewhat analogously, we could prevent much permanent loss of biodiversity by moving more aggressively to preserve genetic material from endangered species in biobanks. The Norwegian government has recently opened a seed bank on a remote island in the arctic archipelago of Svalbard. The vault, which is dug into a mountain and protected by steel-reinforced concrete walls one metre thick, will preserve germplasm of important agricultural and wild plants.

96% of these weapons held by the United States and Russia. Relationships between these two nations are not as bad as they once were. New scares such as environmental problems and terrorism compete effectively for media attention. Changing winds in horror-fashion aside, however, and as Chapter 18 makes it clear, nuclear war remains a very serious threat.

There are several possibilities. One is that relations between the United States and Russia might again worsen to the point where a crisis could trigger a nuclear war. Future arms races could lead to arsenals even larger than those of the past. The world's supply of plutonium has been increasing steadily to about 2000 tons – about ten times as much as remains tied up in warheads – and more could be produced. Some studies suggest that in an all-out war involving most of the weapons in the current US and Russian arsenals, 35–77% of the US population (105–230 million people) and 20–40% of the Russian population (28–56 million people) would be killed. Delayed and indirect effects – such as economic collapse and a possible nuclear winter – could make the final death toll far greater.

Another possibility is that nuclear war might erupt between nuclear powers other than the old Cold War rivals, a risk that is growing as more nations join the nuclear club, especially nations that are embroiled in volatile regional conflicts, such as India and Pakistan, North Korea, and Israel, perhaps to be joined by Iran or others. One concern is that the more nations get the bomb, the harder it might be to prevent further proliferation. The technology and know-how would become more widely disseminated, lowering the technical barriers, and nations that initially chose to forego nuclear weapons might feel compelled to rethink their decision and to follow suit if they see their neighbours start down the nuclear path.

A third possibility is that global nuclear war could be started by mistake. According to Joseph Cirincione, this almost happened in January 1995:

> Russian military officials mistook a Norwegian weather rocket for a US submarine-launched ballistic missile. Boris Yelstin became the first Russian president to ever have the 'nuclear suitcase' open in front of him. He had just a few minutes to decide if he should push the button that would launch a barrage of nuclear missiles. Thankfully, he concluded that his radars were in error. The suitcase was closed.

Several other incidents have been reported in which the world, allegedly, was teetering on the brink of nuclear holocaust. At one point during the Cuban missile crisis, for example, President Kennedy reportedly estimated the probability of a nuclear war between the United States and the USSR to be 'somewhere between one out of three and even'.

To reduce the risks, Cirincione argues, we must work to resolve regional conflicts, support and strengthen the Nuclear Non-proliferation Treaty – one of the most successful security pacts in history – and move towards the abolition of nuclear weapons.

William Potter and Gary Ackerman offer a detailed look at the risks of nuclear terrorism in Chapter 19. Such terrorism could take various forms:

- Dispersal of radioactive material by conventional explosives ('dirty bomb')
- Sabotage of nuclear facilities
- Acquisition of fissile material leading to the fabrication and detonation of a crude nuclear bomb ('improvised nuclear device')
- Acquisition and detonation of an intact nuclear weapon
- The use of some means to trick a nuclear state into launching a nuclear strike.

Potter and Ackerman focus on 'high consequence' nuclear terrorism, which they construe as those involving the last three alternatives from the above list. The authors analyse the demand and supply side of nuclear terrorism, the consequences of a nuclear terrorist attack, the future shape of the threat, and conclude with policy recommendations.

To date, no non-state actor is believed to have gained possession of a fission weapon:

> There is no credible evidence that either al Qaeda or Aum Shinrikyo were able to exploit their high motivations, substantial financial resources, demonstrated organizational skills, far-flung network of followers, and relative security in a friendly or tolerant host country to move very far down the path toward acquiring a nuclear weapons capability. As best one can tell from the limited information available in public sources, among the obstacles that proved most difficult for them to overcome was access to the fissile material needed . . .

Despite this track record, however, many experts remain concerned. Graham Allison, author of one of the most widely cited works on the subject, offers a standing bet of 51 to 49 odds that 'barring radical new anti-proliferation steps' there will be a terrorist nuclear strike within the next ten years. Other experts seem to place the odds much lower, but have apparently not taken up Allison's offer.

There is wide recognition of the importance of prevention nuclear terrorism, and in particular of the need to prevent fissile material from falling into the wrong hands. In 2002, the G-8 Global Partnership set a target of 20 billion dollars to be committed over a ten-year period for the purpose of preventing terrorists from acquiring weapons and materials of mass destruction. What Potter and Ackerman consider most lacking, however, is the sustained high-level leadership needed to transform rhetoric into effective implementation.

In Chapter 20, Christopher Chyba and Ali Nouri review issues related to biotechnology and biosecurity. While in some ways paralleling nuclear risks – biological as well as nuclear technology can be used to build weapons of mass destruction – there are also important divergences. One difference is that biological weapons can be developed in small, easily concealed facilities and require no unusual raw materials for their manufacture. Another difference is that an infectious biological agent can spread far beyond the site of its original release, potentially across the entire world.

Biosecurity threats fall into several categories, including naturally occurring diseases, illicit state biological weapons programmes, non-state actors and bio-hackers, and laboratory accidents or other inadvertent release of disease agents. It is worth bearing in mind that the number of people who have died in recent years from threats in the first of these categories (naturally occurring diseases) is six or seven orders of magnitudes larger than the number of fatalities from the other three categories combined. Yet biotechnology does contain brewing threats which look set to expand dramatically over the coming years as capabilities advance and proliferate. Consider the following sample of recent developments:

- A group of Australian researchers, looking for ways of controlling the country's rabbit population, added the gene for interleukin-4 to a mousepox virus, hoping thereby to render the animals sterile. Unexpectedly, the virus inhibited the host's immune system and all the animals died, including individuals who had previously been vaccinated. Follow-up work by another group produced a version of the virus that was 100% lethal in vaccinated mice despite the antiviral medication given to the animals.
- The polio virus has been synthesized from readily purchased chemical supplies. When this was first done, it required a protracted cutting-edge research project. Since then, the time needed to synthesize a virus genome comparable in size to the polio virus has been reduced to weeks. The virus that caused the Spanish flu pandemic, which was previously extinct, has also been resynthesized and now exists in laboratories in the United States and in Canada.
- The technology to alter the properties of viruses and other microorganisms is advancing at a rapid pace. The recently developed method of RNA interference provides researchers with a ready means of turning off selected genes in humans and other organisms. 'Synthetic biology' is being established as new field, whose goal is to enable the creation of small biological devices and ultimately new types of microbes.

Reading this list, while bearing in mind that the complete genomes from hundreds of bacteria, fungi, viruses – including Ebola, Marburg, smallpox, and the 1918 Spanish influenza virus – have been sequenced and deposited in a public online database, it is not difficult to concoct in one's imagination frightening possibilities. The technological barriers to the production of super bugs are being steadily lowered even as the biotechnological know-how and equipment diffuse ever more widely.

The dual-use nature of the necessary equipment and expertise, and the fact that facilities could be small and easily concealed, pose difficult challenges for would-be regulators. For any regulatory regime to work, it would also have to strike a difficult balance between prevention of abuses and enablement of research needed to develop treatments and diagnostics (or to obtain other medical or economic benefits). Chyba and Nouri discuss several strategies for promoting biosecurity, including automated review of gene sequences submitted for DNA-synthesizing at centralized facilities. It is likely that biosecurity will grow in importance and that a multipronged approach will be needed to address the dangers from designer pathogens.

Chris Phoenix and Mike Treder (Chapter 21) discuss nanotechnology as a source of global catastrophic risks. They distinguish between 'nanoscale technologies', of which many exist today and many more are in development, and 'molecular manufacturing', which remains a hypothetical future technology (often associated with the person who first envisaged it in detail, K. Eric Drexler). Nanoscale technologies, they argue, appear to pose no new global catastrophic risks, although such technologies could in some cases either augment or help mitigate some of the other risks considered in this volume. Phoenix and Treder consequently devote the bulk of their chapter to considering the capabilities and threats from molecular manufacturing. As with superintelligence, the *present* risk is virtually zero since the technology in question does not yet exist; yet the future risk could be extremely severe.

Molecular nanotechnology would greatly expand control over the structure of matter. Molecular machine systems would enable fast and inexpensive manufacture of microscopic and macroscopic objects built to atomic precision. Such production systems would contain millions of microscopic assembly tools. Working in parallel, these would build objects by adding molecules to a workpiece through positionally controlled chemical reactions. The range of structures that could be built with such technology greatly exceeds that accessible to the biological molecular assemblers (such as ribosome) that exist in nature. Among the things that a nanofactory could build: another nanofactory. A sample of potential applications:

- microscopic nanobots for medical use
- vastly faster computers
- very light and strong diamondoid materials
- new processes for removing pollutants from the environment
- desktop manufacturing plants which can automatically produce a wide range of atomically precise structures from downloadable blueprints
- inexpensive solar collectors
- greatly improved space technology

- mass-produced sensors of many kinds
- weapons, both inexpensively mass-produced and improved conventional weapons, and new kinds of weapons that cannot be built without molecular nanotechnology.

A technology this powerful and versatile could be used for an indefinite number of purposes, both benign and malign.

Phoenix and Treder review a number of global catastrophic risks that could arise with such an advanced manufacturing technology, including war, social and economic disruption, destructive forms of global governance, radical intelligence enhancement, environmental degradation, and 'ecophagy' (small nanobots replicating uncontrollably in the natural environment, consuming or destroying the Earth's biosphere). In conclusion, they offer the following rather alarming assessment:

> In the absence of some type of preventive or protective force, the power of molecular manufacturing products could allow a large number of actors of varying types – including individuals, groups, corporations, and nations – to obtain sufficient capability to destroy all unprotected humans. The likelihood of at least one powerful actor being insane is not small. The likelihood that devastating weapons will be built and released accidentally (possibly through overly sensitive automated systems) is also considerable. Finally, the likelihood of a conflict between two [powers capable of unleashing a mutually assured destruction scenario] escalating until one feels compelled to exercise a doomsday option is also non-zero. This indicates that unless adequate defences can be prepared against weapons intended to be ultimately destructive – a point that urgently needs research – the number of actors trying to possess such weapons must be minimized.

The last chapter of the book, authored by Bryan Caplan, addresses totalitarianism as a global catastrophic risk. The totalitarian governments of Nazi Germany, Soviet Russia, and Maoist China were responsible for tens of millions of deaths in the last century. Compared to a risk like that of asteroid impacts, totalitarianism as a global risk is harder to study in an unbiased manner, and a cross-ideological consensus about how this risk is best to be mitigated is likely to be more elusive. Yet the risks from oppressive forms of government, including totalitarian regimes, must not be ignored. Oppression has been one of the major recurring banes of human development throughout history, it largely remains so today, and it is one to which the humanity remains vulnerable.

As Caplan notes, in addition to being a misfortune in itself, totalitarianism can also amplify other risks. People in totalitarian regimes are often afraid to publish bad news, and the leadership of such regimes is often insulated from criticism and dissenting views. This can make such regimes more likely to overlook looming dangers and to commit serious policy errors (even as evaluated from the standpoint of the self-interest of the rulers). However, as

Caplan notes further, for some types of risk, totalitarian regimes might actually possess an advantage compared to more open and diverse societies. For goals that can be achieved by brute force and massive mobilization of resources, totalitarian methods have often proven effective.

Caplan analyses two factors which he claims have historically limited the durability of totalitarian regimes. The first of these is the problem of succession. A strong leader might maintain a tight grip on power for as long as he lives, but the party faction he represents often stumbles when it comes to appointing a successor that will preserve the status quo, allowing a closet reformer – a sheep in wolf's clothing – to gain the leadership position after a tyrant's death. The other factor is the existence of non-totalitarian countries elsewhere in the world. These provide a vivid illustration to the people living under totalitarianism that things could be much better than they are, fuelling dissatisfaction and unrest. To counter this, leaders might curtail contacts with the external world, creating a 'hermit kingdom' such as Communist Albania or present-day North Korea. However, some information is bound to leak in. Furthermore, if the isolation is too complete, over a period of time, the country is likely to fall far behind economically and militarily, making itself vulnerable to invasion or externally imposed regime change.

It is possible that the vulnerability presented by these two Achilles heels of totalitarianism could be reduced by future developments. Technological advances could help solve the problem of succession. Brain scans might one day be used to screen out closet sceptics within the party. Other forms of novel surveillance technologies could also make it easier to control population. New psychiatric drugs might be developed that could increase docility without noticeably reducing productivity. Life-extension medicine might prolong the lifespan of the leader so that the problem of succession comes up less frequently. As for the existence of non-totalitarian outsiders, Caplan worries about the possible emergence of a world government. Such a government, even if it started out democratic, might at some point degenerate into totalitarianism; and a worldwide totalitarian regime could then have great staying power given its lack of external competitors and alien exemplars of the benefits of political freedom.

To have a productive discussion about matters such as these, it is important to recognize the distinction between two very different stances: 'here a valid consideration in favour of some position *X*' versus '*X* is all-things-considered the position to be adopted'. For instance, as Caplan notes:

> If people lived forever, stable totalitarianism would be a little more likely to emerge, but it would be madness to force everyone to die of old age in order to avert a small risk of being murdered by the secret police in a thousand years.

Likewise, it is possible to favour the strengthening of certain new forms global governance while also recognizing as a legitimate concern the danger of global totalitarianism to which Caplan draws our attention.

1.7 Conclusions and future directions

The most likely global catastrophic risks all seem to arise from human activities, especially industrial civilization and advanced technologies. This is not necessarily an indictment of industry or technology, for these factors deserve much of the credit for creating the values that are now at risk – including most of the people living on the planet today, there being perhaps 30 times more of us than could have been sustained with primitive agricultural methods, and hundreds of times more than could have lived as hunter–gatherers. Moreover, although new global catastrophic risks have been created, many smaller-scale risks have been drastically reduced in many parts of the world, thanks to modern technological society. Local and personal disasters – such as starvation, thirst, predation, disease, and small-scale violence – have historically claimed many more lives than have global cataclysms. The reduction of the aggregate of these smaller-scale hazards may outweigh an increase in global catastrophic risks. To the (incomplete) extent that true risk levels are reflected in actuarial statistics, the world is a safer place than it has ever been: world life expectancy is now sixty-four years, up from fifty in the early twentieth century, thirty-three in Medieval Britain, and an estimated eighteen years during the Bronze Age. Global catastrophic risks are, by definition, the largest in terms of *scope* but not necessarily in terms of their expected severity (probability × harm). Furthermore, technology and complex social organizations offer many important tools for managing the remaining risks. Nevertheless, it is important to recognize that the biggest global catastrophic risks we face today are not purely external; they are, instead, tightly wound up with the direct and indirect, the foreseen and unforeseen, consequences of our own actions.

One major current global catastrophic risk is infectious pandemic disease. As noted earlier, infectious disease causes approximately 15 million deaths per year, of which 75% occur in Southeast Asia and sub-Saharan Africa. These dismal statistics pose a challenge to the classification of pandemic disease as a global catastrophic risk. One could argue that infectious disease is not so much a *risk* as an *ongoing global catastrophe*. Even on a more fine-grained individuation of the hazard, based on specific infectious agents, at least some of the currently occurring pandemics (such as HIV/AIDS, which causes nearly 3 million deaths annually) would presumably qualify as global catastrophes. By similar reckoning, one could argue that cardiovascular disease (responsible for approximately 30% of world mortality, or 18 million deaths per year) and cancer (8 million deaths) are also ongoing global catastrophes. It would be perverse if the study of possible catastrophes that *could* occur were to drain attention away from actual catastrophes that *are* occurring.

It is also appropriate, at this juncture, to reflect for a moment on the biggest cause of death and disability of all, namely ageing, which accounts for perhaps two-thirds of the 57 million deaths that occur each year, along with an enormous loss of health and human capital.[12] If ageing were not certain but merely probable, it would immediately shoot to the top of any list of global catastrophic risks. Yet the fact that ageing is not just a possible cause of future death, but a certain cause of present death, should not trick us into trivializing the matter. To the extent that we have a realistic prospect of mitigating the problem – for example, by disseminating information about healthier lifestyles or by investing more heavily in biogerontological research – we may be able to save a much larger expected numbers of lives (or quality-adjusted life-years) by making partial progress on this problem than by completely eliminating some of the global catastrophic risk discussed in this volume.

Other global catastrophic risks which are either already substantial or expected to become substantial within a decade or so include the risks from nuclear war, biotechnology (misused for terrorism or perhaps war), social/economic disruption or collapse scenarios, and maybe nuclear terrorism. Over a somewhat longer time frame, the risks from molecular manufacturing, artificial intelligence, and totalitarianism may rise in prominence, and each of these latter ones is also potentially existential.

That a particular risk is larger than another does not imply that more resources ought to be devoted to its mitigation. Some risks we might not be able to do anything about. For other risks, the available means of mitigation might be too expensive or too dangerous. Even a small risk can deserve to be tackled as a priority if the solution is sufficiently cheap and easy to implement – one example being the anthropogenic depletion of the ozone layer, a problem now well on its way to being solved. Nevertheless, as a rule of thumb it makes sense to devote most of our attention to the risks that are largest and/or most urgent. A wise person will not spend time installing a burglar alarm when the house is on fire.

Going forward, we need continuing studies of individual risks, particularly of potentially big but still relatively poorly understood risks, such as those from biotechnology, molecular manufacturing, artificial intelligence, and systemic risks (of which totalitarianism is but one instance). We also need studies to identify and evaluate possible mitigation strategies. For some risks and ongoing disasters, cost-effective countermeasures are already known; in these cases, what is needed is leadership to ensure implementation of the appropriate programmes. In addition, there is a need for studies to clarify methodological problems arising in the study of global catastrophic risks.

---

[12] In mortality statistics, deaths are usually classified according to their more proximate causes (cancer, suicide, etc.). But we can estimate how many deaths are due to ageing by comparing the age-specific mortality in different age groups. The reason why an average 80-year-old is more likely to die within the next year than an average 20-year-old is that senescence has made the former more susceptible to a wide range of specific risk factors. The surplus mortality in older cohorts can therefore be attributed to the negative effects of ageing.

The fruitfulness of further work on global catastrophic risk will, we believe, be enhanced if it gives consideration to the following suggestions:

- In the study of individual risks, focus more on producing actionable information such as early-warning signs, metrics for measuring progress towards risk reduction, and quantitative models for risk assessment.
- Develop and implement better methodologies and institutions for information aggregation and probabilistic forecasting, such as prediction markets.
- Put more effort into developing and evaluating possible mitigation strategies, both because of the direct utility of such research and because a concern with the policy instruments with which a risk can be influenced is likely to enrich our theoretical understanding of the nature of the risk.
- Devote special attention to existential risks and the unique methodological problems they pose.
- Build a stronger interdisciplinary and international risk community, including not only experts from many parts of academia but also professionals and policymakers responsible for implementing risk reduction strategies, in order to break out of disciplinary silos and to reduce the gap between theory and practice.
- Foster a critical discourse aimed at addressing questions of prioritization in a more reflective and analytical manner than is currently done; and consider global catastrophic risks and their mitigation within a broader context of challenges and opportunities for safeguarding and improving the human condition.

Our hopes for this book will have been realized if it adds a brick to the foundation of a way of thinking that enables humanity to approach the global problems of the present era with greater maturity, responsibility, and effectiveness.

This page intentionally left blank

PART I

Background

This page intentionally left blank

·2·

Long-term astrophysical processes

Fred C. Adams

2.1 Introduction: physical eschatology

As we take a longer-term view of our future, a host of astrophysical processes are waiting to unfold as the Earth, the Sun, the Galaxy, and the Universe grow increasingly older. The basic astronomical parameters that describe our universe have now been measured with compelling precision. Recent observations of the cosmic microwave background radiation show that the spatial geometry of our universe is flat (Spergel et al., 2003). Independent measurements of the red-shift versus distance relation using Type Ia supernovae indicate that the universe is accelerating and apparently contains a substantial component of dark vacuum energy (Garnavich et al., 1998; Perlmutter et al., 1999; Riess et al., 1998).[1] This newly consolidated cosmological model represents an important milestone in our understanding of the cosmos. With the cosmological parameters relatively well known, the future evolution of our universe can now be predicted with some degree of confidence (Adams and Laughlin, 1997). Our best astronomical data imply that our universe will expand forever or at least live long enough for a diverse collection of astronomical events to play themselves out.

Other chapters in this book have discussed some sources of cosmic intervention that can affect life on our planet, including asteroid and comet impacts (Chapter 11, this volume) and nearby supernova explosions with their accompanying gamma-rays (Chapter 12, this volume). In the longer-term future, the chances of these types of catastrophic events will increase. In addition, taking an even longer-term view, we find that even more fantastic events could happen in our cosmological future. This chapter outlines some of the astrophysical events that can affect life, on our planet and perhaps

---

[1] 'Dark energy' is a common term unifying different models for the ubiquitous form of energy permeating the entire universe (about 70% of the total energy budget of the physical universe) and causing accelerated expansion of space time. The most famous of these models is Einstein's *cosmological constant*, but there are others, going under the names of *quintessence*, *phantom energy*, and so on. They are all characterized by negative pressure, in sharp contrast to all other forms of energy we see around us.

elsewhere, over extremely long time scales, including those that vastly exceed the current age of the universe.

These projections are based on our current understanding of astronomy and the laws of physics, which offer a firm and developing framework for understanding the future of the physical universe (this topic is sometimes called *Physical Eschatology* – see the review of Ćirković, 2003). Notice that as we delve deeper into the future, the uncertainties of our projections must necessarily grow. Notice also that this discussion is based on the assumption that the laws of physics are both known and unchanging; as new physics is discovered, or if the physical constants are found to be time dependent, this projection into the future must be revised accordingly.

2.2 Fate of the Earth

One issue of immediate importance is the fate of Earth's biosphere and, on even longer time scales, the fate of the planet itself. As the Sun grows older, it burns hydrogen into helium. Compared to hydrogen, helium has a smaller partial pressure for a given temperature, so the central stellar core must grow hotter as the Sun evolves. As a result, the Sun, like all stars, is destined to grow brighter as it ages. When the Sun becomes too bright, it will drive a runaway greenhouse effect through the Earth's atmosphere (Kasting et al., 1988). This effect is roughly analogous to that of global warming driven by greenhouse gases (see Chapter 13, this volume), a peril that our planet faces in the near future; however, this later-term greenhouse effect will be much more severe. Current estimates indicate that our biosphere will be essentially sterilized in about 3.5 billion years, so this future time marks the end of life on Earth. The end of *complex* life may come sooner, in 0.9–1.5 billion years owing to the runaway greenhouse effect (e.g., Caldeira and Kasting, 1992).

The biosphere represents a relatively small surface layer and the planet itself lives comfortably through this time of destruction. Somewhat later in the Sun's evolution, when its age reaches 11–12 billion years, it eventually depletes its store of hydrogen in the core region and must readjust its structure (Rybicki and Denis, 2001; Sackmann et al., 1993). As it does so, the outer surface of the star becomes somewhat cooler, its colour becomes a brilliant red, and its radius increases. The red giant Sun eventually grows large enough to engulf the radius of the orbit of Mercury, and that innermost planet is swallowed with barely a trace left. The Sun grows further, overtakes the orbit of Venus, and then accretes the second planet as well. As the red giant Sun expands, it loses mass so that surviving planets are held less tightly in their orbits. Earth is able to slip out to an orbit of larger radius and seemingly escape destruction. However, the mass loss from the Sun provides a fluid that the Earth must plough through as it makes its yearly orbit. Current calculations indicate that the frictional forces acting on Earth through its interaction with the solar outflow cause the planet to experience enough orbital decay that it is dragged back into the Sun. Earth is thus evaporated, with its legacy being a small addition to the heavy element supply of the solar photosphere. This point in future history, approximately 7 billion years from now, marks the end of our planet.

Given that the biosphere has at most only 3.5 billion years left on its schedule, and Earth itself has only 7 billion years, it is interesting to ask what types of 'planet-saving' events can take place on comparable time scales. Although the odds are not good, the Earth has some chance of being 'saved' by being scattered out of the solar system by a passing star system (most of which are binary stars). These types of scattering interactions pose an interesting problem in solar system dynamics, one that can be addressed with numerical scattering experiments. A large number of such experiments must be run because the systems are chaotic, and hence display sensitive dependence on their initial conditions, and because the available parameter space is large. Nonetheless, after approximately a half million scattering calculations, an answer can be found: the odds of Earth being ejected from the solar system before it is accreted by the red giant Sun is a few parts in 10^5 (Laughlin and Adams, 2000).

Although sending the Earth into exile would save the planet from eventual evaporation, the biosphere would still be destroyed. The oceans would freeze within a few million years and the only pockets of liquid water left would be those deep underground. The Earth contains an internal energy source – the power produced by the radioactive decay of unstable nuclei. This power is about 10,000 times smaller than the power that Earth intercepts from the present-day Sun, so it has little effect on the current operation of the surface biosphere. If Earth were scattered out of the solar system, then this internal power source would be the only one remaining. This power is sufficient to keep the interior of the planet hot enough for water to exist in liquid form, but only at depths 14 km below the surface. This finding, in turn, has implications for present-day astronomy: the most common liquid water environments may be those deep within frozen planets, that is, those that have frozen water on their surfaces and harbour oceans of liquid water below. Such planets may be more common than those that have water on their surface, like Earth, because they can be found in a much wider range of orbits about their central stars (Laughlin and Adams, 2000).

In addition to saving the Earth by scattering it out of the solar system, passing binaries can also capture the Earth and thereby allow it to orbit about a new star. Since most stars are smaller in mass than our Sun, they live longer and suffer less extreme red giant phases. (In fact, the smallest stars with less than one-fourth of the mass of the Sun will never become red giants – Laughlin et al., 1997.) As a result, a captured Earth would stand a better chance of long-term survival. The odds for this type of planet-saving event taking place while the biosphere remains intact are exceedingly slim – only about one in three million (Laughlin and Adams, 2000), roughly the odds of winning a big state lottery.

For completeness, we note that in addition to the purely natural processes discussed here, human or other intentional intervention could potentially change the course of Earth's orbit given enough time and other resources. As a concrete example, one could steer an asteroid into the proper orbit so that gravitational scattering effectively transfers energy into the Earth's orbit, thereby allowing it to move outward as the Sun grows brighter (Korycansky et al., 2001). In this scenario, the orbit of the asteroid is chosen to encounter both Jupiter and Saturn, and thereby regain the energy and angular momentum that it transfers to Earth. Many other scenarios are possible, but the rest of this chapter will focus on physical phenomena not including intentional actions.

2.3 Isolation of the local group

Because the expansion rate of the universe is starting to accelerate (Garnavich et al., 1998; Perlmutter et al., 1999; Riess et al., 1998), the formation of galaxies, clusters, and larger cosmic structures is essentially complete. The universe is currently approaching a state of exponential expansion and growing cosmological fluctuations will freeze out on all scales. Existing structures will grow isolated. Numerical simulations illustrate this trend (Fig. 2.1) and show how the universe will break up into a collection of 'island universes', each containing one bound cluster or group of galaxies (Busha et al., 2003; Nagamine and Loeb, 2003). In other words, the largest gravitationally bound structures that we see in the universe today are likely to be the largest structures that ever form. Not only must each group of galaxies (eventually) evolve in physical isolation, but the relentless cosmic expansion will stretch existing galaxy clusters out of each others' view. In the future, one will not even be able to see the light from galaxies living in other clusters. In the case of the Milky Way, only the Local Group of Galaxies will be visible. Current observations and recent numerical studies clearly indicate that the nearest large cluster – Virgo – does not have enough mass for the Local Group to remain bound to it in the future (Busha et al., 2003; Nagamine and Loeb, 2003). This local group consists of the Milky Way, Andromeda, and a couple of dozen dwarf galaxies (irregulars and spheroidals). The rest of the universe will be cloaked behind a cosmological horizon and hence will be inaccessible to future observation.

2.4 Collision with Andromeda

Within their clusters, galaxies often pass near each other and distort each other's structure with their strong gravitational fields. Sometimes these

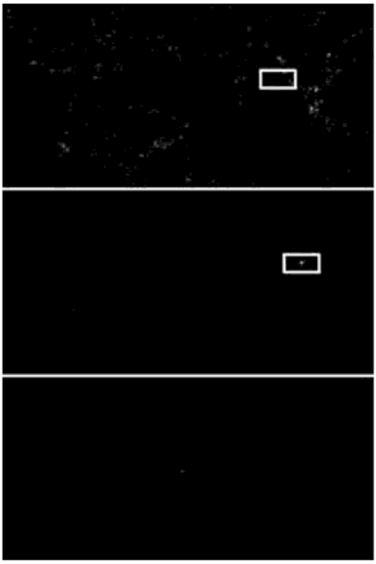

Fig. 2.1 Numerical simulation of structure formation in an accelerating universe with dark vacuum energy. The top panel shows a portion of the universe at the present time (cosmic age 14 Gyr). The boxed region in the upper panel expands to become the picture in the central panel at cosmic age 54 Gyr. The box in the central panel then expands to become the picture shown in the bottom panel at cosmic age 92 Gyr. At this future epoch, the galaxy shown in the centre of the bottom panel has grown effectively isolated. (Simulations reprinted with permission from Busha, M.T., Adams, F.C., Evrard, A.E., and Wechsler, R.H. (2003). Future evolution of cosmic structure in an accelerating universe. *Astrophys. J.*, 596, 713.)

interactions lead to galactic collisions and merging. A rather important example of such a collision is coming up: the nearby Andromeda galaxy is headed straight for our Milky Way. Although this date with our sister galaxy will not take place for another 6 billion years or more, our fate is sealed – the two galaxies are a bound pair and will eventually merge into one (Peebles, 1994).

When viewed from the outside, galactic collisions are dramatic and result in the destruction of the well-defined spiral structure that characterizes the original galaxies. When viewed from within the galaxy, however, galactic collisions are considerably less spectacular. The spaces between stars are so vast that few, if any, stellar collisions take place. One result is the gradual brightening of the night sky, by roughly a factor of two. On the other hand, galactic collisions are frequently associated with powerful bursts of star formation. Large clouds of molecular gas within the galaxies merge during such collisions and produce new stars at prodigious rates. The multiple supernovae resulting from the deaths of the most massive stars can have catastrophic consequences and represent a significant risk to any nearby biosphere (see Chapter 12, this volume), provided that life continues to thrive in thin spherical layers on terrestrial planets.

2.5 The end of stellar evolution

With its current age of 14 billion years, the universe now lives in the midst of a Stelliferous Era, an epoch when stars are actively forming, living, and dying. Most of the energy generated in our universe today arises from nuclear fusion that takes place in the cores of ordinary stars. As the future unfolds, the most common stars in the universe – the low-mass stars known as red dwarfs – play an increasingly important role. Although red dwarf stars have less than half the mass of the Sun, they are so numerous that their combined mass easily dominates the stellar mass budget of the galaxy. These red dwarfs are parsimonious when it comes to fusing their hydrogen into helium. By hoarding their energy resources, they will still be shining trillions of years from now, long after their larger brethren have exhausted their fuel and evolved into white dwarfs or exploded as supernovae. It has been known for a long time that smaller stars live much longer than more massive ones owing to their much smaller luminosities. However, recent calculations show that red dwarfs live even longer than expected. In these small stars, convection currents cycle essentially all of the hydrogen fuel in the star through the stellar core, where it can be used as nuclear fuel. In contrast, our Sun has access to only about 10% of its hydrogen and will burn only 10% of its nuclear fuel while on the main sequence. A small star with 10% of the mass of the Sun thus has nearly the same fuel reserves and will shine for tens of trillions of years (Laughlin et al., 1997). Like all stars, red dwarfs get brighter as they age. Owing to their large population, the brightening of red dwarfs nearly compensates for the loss of larger stars, and the galaxy can maintain a nearly constant luminosity for approximately one trillion years (Adams et al., 2004).

Even small stars cannot live forever, and this bright stellar era comes to a close when the galaxies run out of hydrogen gas, star formation ceases, and the longest-lived red dwarfs slowly fade into oblivion. As mentioned earlier, the smallest stars will shine for trillions of years, so the era of stars would come to an end at a cosmic age of several trillion years if new stars were not being manufactured. In large spiral galaxies like the Milky Way, new stars are being made from hydrogen gas, which represents the basic raw material for the process. Galaxies will continue to make new stars as long as the gas supply holds out. If our Galaxy were to continue forming stars at its current rate, it would run out of gas in 'only' 10–20 billion years (Kennicutt et al., 1994), much shorter than the lifetime of the smallest stars. Through conservation practices – the star formation rate decreases as the gas supply grows smaller – galaxies can sustain normal star formation for almost the lifetime of the longest-lived stars (Adams and Laughlin, 1997; Kennicutt et al., 1994). Thus, both stellar evolution and star formation will come to an end at approximately the same time in our cosmic future. The universe will be about 100 trillion (10^{14}) years old when the stars finally stop shining. Although our Sun will have long since burned out, this time marks an important turning point for any surviving biospheres – the power available is markedly reduced after the stars turn off.

2.6 The era of degenerate remnants

After the stars burn out and star formation shuts down, a significant fraction of the ordinary mass will be bound within the degenerate remnants that remain after stellar evolution has run its course. For completeness, however, one should keep in mind that the majority of the baryonic matter will remain in the form of hot gas between galaxies in large clusters (Nagamine and Loeb, 2004). At this future time, the inventory of degenerate objects includes brown dwarfs, white dwarfs, and neutron stars. In this context, degeneracy refers to the state of the high-density material locked up in the stellar remnants. At such enormous densities, the quantum mechanical exclusion principle determines the pressure forces that hold up the stars. For example, when most stars die, their cores shrink to roughly the radial size of Earth. With this size, the density of stellar material is about one million times greater than that of the Sun, and the pressure produced by degenerate electrons holds up the star against further collapse. Such objects are white dwarfs and they will contain most of the mass in stellar bodies at this epoch. Some additional mass is contained in brown dwarfs, which are essentially failed stars that never fuse hydrogen, again owing to the effects of degeneracy pressure. The largest stars, those that begin with masses more than eight times that of the Sun, explode at the end of their lives as supernovae. After the explosion, the stellar cores are compressed to densities about one quadrillion times that of the Sun. The resulting stellar body is a neutron star, which is held up by the degeneracy pressure of its constituent neutrons (at such enormous densities, typically *a few* $\times 10^{15}$ g/cm^3, electrons and protons combine to form neutrons, which make the star much like a gigantic atomic nucleus). Since only three or four out of every thousand stars are massive enough to produce a supernova explosion, neutron stars will be rare objects.

During this Degenerate Era, the universe will look markedly different from the way it appears now. No visible radiation from ordinary stars will light up the skies, warm the planets, or endow the galaxies with the faint glow they have today. The cosmos will be darker, colder, and more desolate. Against this stark backdrop, events of astronomical interest will slowly take place. As dead stars trace through their orbits, close encounters lead to scattering events, which force the galaxy to gradually readjust its structure. Some stellar remnants are ejected beyond the reaches of the galaxy, whereas others fall in toward the centre. Over the next 10^{20} years, these interactions will enforce the dynamical destruction of the entire galaxy (e.g., Binney and Tremaine, 1987; Dyson, 1979).

In the meantime, brown dwarfs will collide and merge to create new low-mass stars. Stellar collisions are rare because the galaxy is relentlessly empty. During this future epoch, however, the universe will be old enough so that some collisions will occur, and the merger products will often be massive enough to sustain hydrogen fusion. The resulting low-mass stars will then burn for trillions of years. At any given time, a galaxy the size of our Milky Way will harbour a few stars formed through this unconventional channel (compare this stellar population with the approximately 100 billion stars in the Galaxy today).

Along with the brown dwarfs, white dwarfs will also collide at roughly the same rate. Most of the time, such collisions will result in somewhat larger white dwarfs. More rarely, white dwarf collisions produce a merger product with a mass greater than the Chandrasekhar limit. These objects will result in a supernova explosion, which will provide spectacular pyrotechnics against the dark background of the future galaxy.

White dwarfs will contain much of the ordinary baryonic matter in this future era. In addition, these white dwarfs will slowly accumulate weakly interacting dark matter particles that orbit the galaxy in an enormous diffuse halo. Once trapped within the interior of a white dwarf, the particles annihilate each other and provide an important source of energy for the cosmos. Dark matter annihilation will replace conventional nuclear burning in stars as the dominant energy source. The power produced by this process is much lower than that produced by nuclear burning in conventional stars. White dwarfs fuelled by dark matter annihilation produce power ratings measured in quadrillions of Watts, roughly comparable to the total solar power intercepted by Earth (approximately 10^{17} Watts). Eventually, however, white dwarfs will be ejected from the galaxy, the supply of the dark matter will get depleted, and this method of energy generation must come to an end.

Although the proton lifetime remains uncertain, elementary physical considerations suggest that protons will not live forever. Current experiments show that the proton lifetime is longer than about 10^{33} years (Super-Kamiokande Collaboration, 1999), and theoretical arguments (Adams et al., 1998; Ellis et al., 1983; Hawking et al., 1979; Page, 1980; Zeldovich, 1976) suggest that the proton lifetime should be less than about 10^{45} years. Although this allowed range of time scales is rather large, the mass-energy stored within white dwarfs and other degenerate remnants will eventually evaporate when their constituent protons and neutrons decay. As protons decay inside a white dwarf, the star generates power at a rate that depends on the proton lifetime. For a value near the centre of the (large) range of allowed time scales (specifically 10^{37} years), proton decay within a white dwarf generates approximately 400 Watts of power – enough to run a few light bulbs. An entire galaxy of these stars will appear dimmer than our present-day Sun. The process of proton decay converts the mass energy of the particles into radiation, so the white dwarfs evaporate away. As the proton decay process grinds to completion, perhaps 10^{40} years from now, all of the degenerate stellar remnants disappear from the universe. This milestone marks a definitive end to life as we know it, as no carbon-based life can survive the cosmic catastrophe induced by proton decay. Nonetheless, the universe continues to exist, and astrophysical processes continue beyond this end of known biology.

2.7 The era of black holes

After the protons decay, the universe grows even darker and more rarefied. At this late time, roughly when the universe is older than 10^{45} years, the only stellar-like objects remaining are black holes. They are unaffected by proton decay and slide unscathed through the end of the previous era. These objects are often defined to be regions of space-time with such strong gravitational fields that even light cannot escape from their surfaces. But at this late epoch, black holes will be the brightest objects in the sky. Thus, even black holes cannot last forever. They shine ever so faintly by emitting a nearly thermal spectrum of photons, gravitons, and other particles (Hawking, 1974). Through this quantum mechanical process – known as Hawking radiation – black holes convert their mass into radiation and evaporate at a glacial pace (Fig. 2.2). In the far future, black holes will provide the universe with its primary source of power.

Although their energy production via Hawking radiation will not become important for a long time, the production of black holes, and hence the black hole inventory of the future, is set by present-day (and past) astrophysical processes. Every large galaxy can produce millions of stellar black holes, which result from the death of the most massive stars. Once formed, these black holes will endure for up to 10^{70} years. In addition, almost every galaxy harbours a super-massive black hole anchoring its centre; these monsters were produced during the process of galaxy formation, when the universe was only a billion years old, or perhaps even younger. They gain additional mass with time and provide the present-day universe with accretion power. As these large black holes evaporate through the Hawking process, they can last up to 10^{100} years. But even the largest black holes must ultimately evaporate. This Black Hole Era will be over when the largest black holes have made their explosive exits from our universe.

2.8 The Dark Era and beyond

When the cosmic age exceeds 10^{100} years, the black holes will be gone and the cosmos will be filled with the leftover waste products from previous

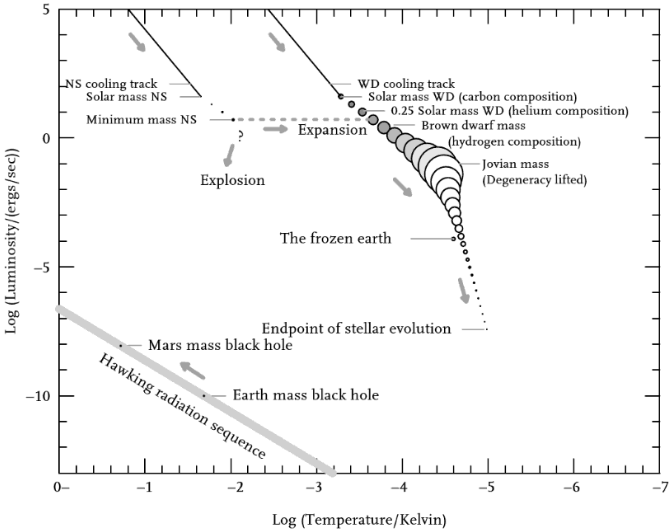

Fig. 2.2 This plot shows the long-term evolution of cold degenerate stars in the H-R diagram. After completing the early stages of stellar evolution, white dwarfs and neutron stars cool to an equilibrium temperature determined by proton decay. This figure assumes that proton decay is driven by gravity (microscopic black holes) on a time scale of 10^{45} years. The white dwarf models are plotted at successive twofold decrements in mass. The mean stellar density (in $\log[\rho/g]$) is indicated by the grey scale shading, and the sizes of the circles are proportional to stellar radius. The relative size of the Earth and its position on the diagram are shown for comparison. The evaporation of a neutron star, starting with one solar mass, is illustrated by the parallel sequence, which shows the apparent radial sizes greatly magnified for clarity. The Hawking radiation sequence for black holes is also plotted. The arrows indicate the direction of time evolution. (Reprinted with permission from Adams, F.C., Laughlin, G., Mbonye, M., and Perry, M.J. (1998). Gravitational demise of cold degenerate stars. *Phys. Rev. D*, 58, 083003.)

eras: neutrinos, electrons, positrons, dark matter particles, and photons of incredible wavelength. In this cold and distant Dark Era, physical activity in the universe slows down, almost (but not quite) to a standstill. The available energy is limited and the expanses of time are staggering, but the universe doggedly continues to operate. Chance encounters between electrons and positrons can forge positronium atoms, which are exceedingly rare in an accelerating universe. In addition, such atoms are unstable and eventually decay. Other low-level annihilation events also take place, for example, between any surviving dark matter particles. In the poverty of this distant epoch, the generation of energy and entropy becomes increasingly difficult.

At this point in the far future, predictions of the physical universe begin to lose focus. If we adopt a greater tolerance for speculation, however, a number of possible events can be considered. One of the most significant potential events is that the vacuum state of the universe could experience a phase transition to a lower energy state. Our present-day universe is observed to be accelerating, and one possible implication of this behaviour is that empty space has a non-zero energy associated with it. In other words, empty space is not really empty, but rather contains a positive value of vacuum energy. If empty space is allowed to have a non-zero energy (allowed by current theories of particle physics), then it remains possible for empty space to have two (or more) different accessible energy levels. In this latter case, the universe could make a transition from its current (high energy) vacuum state to a lower-energy state sometime in the future (the possibility of inducing such a phase transition is discussed in Chapter 16). As the universe grows increasingly older, the probability of a spontaneous transition grows as well. Unfortunately, our current understanding of the vacuum state of the universe is insufficient to make a clear predictions on this issue – the time scale for the transition remains enormously uncertain. Nonetheless, such a phase transition remains an intriguing possibility. If the universe were to experience a vacuum phase transition, it remains possible (but is not guaranteed) that specific aspects of the laws of physics (e.g., the masses of the particles and/or the strengths of the forces) could change, thereby giving the universe a chance for a fresh start.

2.9 Life and information processing

The discussion in this chapter has focused on physical processes that can take place in the far future. But what about life? How far into the future can living organisms survive? Although this question is of fundamental importance and holds enormous interest, our current understanding of biology is not sufficiently well developed to provide a clear answer. To further complicate matters, protons must eventually decay, as outlined above, so that carbon-based life will come to a definitive end. Nonetheless, some basic principles can be discussed if we are willing to take a generalized view of life, where we consider life to be essentially a matter of information processing. This point of view has been pioneered by Freeman Dyson (1979), who argued that the rate of metabolism or information processing in a generalized life form should be proportional to its operating temperature.

If our universe is accelerating, as current observations indicate, then the amount of matter and hence energy accessible to a given universe will be finite. If the operating temperature of life remains constant, then this finite free energy would eventually be used up and life would come to an end. The only chance for continued survival is to make the operating temperature of life decrease. More specifically, the temperature must decrease fast enough to allow for an infinite amount of information processing with a finite amount of free energy.

According to the Dyson scaling hypothesis, as the temperature decreases, the rate of information processing decreases, and the quality of life decreases accordingly. Various strategies to deal with this problem have been discussed, including the issue of digital versus analogous life, maintaining long-term survival by long dormant periods (hibernation), and the question of classical versus quantum mechanical information processing (e.g., Dyson, 1979; Krauss and Starkman, 2000). Although a definitive conclusion has not been reached, the prospects are rather bleak for the continued (infinite) survival of life. The largest hurdle seems to be continued cosmic acceleration, which acts to limit the supply of free energy. If the current acceleration comes to an end, so that the future universe expands more slowly, then life will have a better chance for long-term survival.

2.10 Conclusion

As framed by a well-known poem by Robert Frost, the world could end either in fire or in ice. In the astronomical context considered here, Earth has only a small chance of escaping the fiery wrath of the red giant Sun by becoming dislodged from its orbit and thrown out into the icy desolation of deep space. Our particular world is thus likely to end its life in fire. Given that humanity has a few billion years to anticipate this eventuality, one can hope that migration into space could occur, provided that the existential disasters outlined in other chapters of this book can be avoided. One alternative is for a passing star to wander near the inner portion of our solar system. In this unlikely event, the disruptive gravitational effects of the close encounter could force Earth to abandon its orbit and be exiled from the solar system. In this case, our world would avoid a scalding demise, but would face a frozen future.

A similar fate lies in store for the Sun, the Galaxy, and the Universe. At the end of its life as an ordinary star, the Sun is scheduled to become a white dwarf. This stellar remnant will grow increasingly cold and its nuclei will atrophy to lower atomic numbers as the constituent protons decay. In the long run, the Sun will end up as a small block of hydrogen ice. As it faces its demise, our Galaxy will gradually evaporate, scattering its stellar bodies far and wide. The effective temperature of a stellar system is given by the energies of its stellar orbits. In the long term, these energies will fade to zero and the galaxy will end its life in a cold state. For the universe as a whole, the future is equally bleak, but far more drawn out. The currently available astronomical data indicate that the universe will expand forever, or at least for long enough that the timeline outlined above can play itself out. As a result, the cosmos, considered as a whole, is likely to grow ever colder and face an icy death.

In the beginning, starting roughly fourteen billion years ago, the early universe consisted of elementary particles and radiation – essentially because the background was too hot for larger structures to exist. Here we find that the universe of the far future will also consist of elementary particles and radiation – in this case because the cosmos will be too cold for larger entities to remain intact. From this grand perspective, the galaxies, stars, and planets that populate the universe today are but transient phenomena, destined to fade into the shifting sands of time. Stellar remnants, including the seemingly resilient black holes, are also scheduled to decay. Even particles as fundamental as protons will not last forever. Ashes to ashes, dust to dust, particles to particles – such is the ultimate fate of our universe.

Suggestions for further reading

Adams, F.C. and Laughlin, G. (1997). A dying universe: the long term fate and evolution of astrophysical objects. *Rev. Mod. Phys.*, **69**, pp. 337–372. This review outlines the physics of the long-term future of the universe and its constituent astrophysical objects (advanced level).

Adams, F.C. and Laughlin, G. (1999). *Five Ages of the Universe: Inside the Physics of Eternity* (New York: The Free Press). Provides a popular level account of the future history of the universe.

Ćirković, M.M. (2003). Resource letter Pes-1: physical eschatology. *Am. J. Phys.*, **71**, pp. 122–133. This paper provides a comprehensive overview of the scientific literature concerning the future of the universe (as of 2003). The treatment is broad and also includes books, popular treatments, and philosophical accounts (advanced level).

Dyson, F.J. (1979). Time without end: physics and biology in an open universe. *Rev. Mod. Phys.*, **51**, pp. 447–460. This review represents one of the first comprehensive treatments of the future of the universe and includes discussion of the future of both communication and biology (advanced level).

Islam, J.N. (1983). *The Ultimate Fate of the Universe* (Cambridge: Cambridge University Press). This book provides one of the first popular level accounts of the future of the universe and raises for the first time many subsequently discussed questions.

Rees, M. (1997). *Before the Beginning: Our Universe and Others* (Reading, MA: Addision-Wesley). This book provides a popular level treatment of the birth of the universe and hence the starting point for discussions of our cosmic future.

References

Adams, F.C. and Laughlin, G. (1997). A dying universe: the long term fate and evolution of astrophysical objects. *Rev. Mod. Phys.*, **69**, 337–372.

Adams, F.C., Laughlin, G., and Graves, G.J.M. (2004). Red dwarfs and the end of the main sequence. *Rev. Mexican Astron. Astrophys.*, **22**, 46–49.

Adams, F.C., Laughlin, G., Mbonye, M., and Perry, M.J. (1998). Gravitational demise of cold degenerate stars. *Phys. Rev. D*, **58**, 083003 (7 pages).

Binney, J. and Tremaine, S. (1987). *Galactic Dynamics* (Princeton, NJ: Princeton University Press).

Busha, M.T., Adams, F.C., Evrard, A.E., and Wechsler, R.H. (2003). Future evolution of cosmic structure in an accelerating universe. *Astrophys. J.*, **596**, 713–724.

Caldeira, K. and Kasting, J.F. (1992). The life span of the biosphere revisited. *Nature*, **360**, 721–723.

Ćirković, M.M. (2003). Resource letter PEs-1: physical eschatology. *Am. J. Phys.*, **71**, 122–133.

Dyson, F.J. (1979). Time without end: physics and biology in an open universe. *Rev. Mod. Phys.*, **51**, 447–460.

Ellis, J., Hagelin, J.S., Nanopoulos, D.V., and Tamvakis, K. (1983). Observable gravitationally induced baryon decay. *Phys. Lett.*, **B**124, 484–490.

Garnavich, P.M., Jha, S., Challis, P., Clocchiatti, A., Diercks, A., Filippenko, A.V., Gilliland, R.L., Hogan, C.J., Kirshner, R.P., Leibundgut, B., Phillips, M.M., Reiss, D., Riess, A.G., Schmidt, B.P., Schommer, R.A., Smith, R.C., Spyromilio, J., Stubbs, C., Suntzeff, N.B., Tonry, J., and Carroll, S.M. (1998). Supernova limits on the cosmic equation of state. *Astrophys. J.*, **509**, 74–79.

Hawking, S.W. (1974). Black hole explosions? *Nature*, **248**, 30–31.

Hawking, S.W., Page, D.N., and Pope, C.N. (1979). The propagation of particles in space-time foam. *Phys. Lett.*, **B**86, 175–178.

Kasting, J.F. (1988). Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. *Icarus*, **74**, 472–494.

Kennicutt, R.C., Tamblyn, P., and Congdon, C.W. (1994). Past and future star formation in disk galaxies. *Astrophys. J.*, **435**, 22–36.

Korycansky, D.G., Laughlin, G., and Adams, F.C. (2001). Astronomical Engineering: a strategy for modifying planetary orbits. *Astrophys. Space Sci.*, **275**, 349–366.

Krauss, L.M. and Starkman, G.D. (2000). Life, the Universe, and Nothing: life and death in an ever-expanding universe. *Astrophys. J.*, **531**, 22–30.

Laughlin, G. and Adams, F.C. (2000). The frozen Earth: binary scattering events and the fate of the Solar System. *Icarus*, **145**, 614–627.

Laughlin, G., Bodenheiemr, P., and Adams, F.C. (1997). The end of the main sequence. *Astrophys. J.*, **482**, 420.

Nagamine, K. and Loeb, A. (2003). Future evolution of nearby large-scale structures in a universe dominated by a cosmological constant. *New Astron.*, **8**, 439–448.

Nagamine, K. and Loeb, A. (2004). Future evolution of the intergalactic medium in a universe dominated by a cosmological constant. *New Astron.*, **9**, 573–583.

Page, D. N. (1980). Particle transmutations in quantum gravity. *Phys. Lett.*, **B**95, 244–246.

Peebles, P.J.E. (1994). Orbits of the nearby galaxies. *Astrophys. J.*, **429**, 43–65.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R.A., Nugent, P., Castro, P.G., Deustua, S., Fabbro, S., Goobar, A., Groom, D.E., Hook, I.M., Kim, A.G., Kim, M.Y., Lee, J.C., Nunes, N.J., Pain, R., Pennypacker, C.R., Quimby, R., Lidman, C., Ellis, R.S., Irwin, M., McMahon, R.G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B.J., Filippenko, A.V., Matheson, T., Fruchter, A.S., Panagia, N., Newberg, H.J.M., and Couch, W.J. (1999). Measurements of Ω and Λ from 42 high-redshift supernovae. *Astrophys. J.*, **517**, 565–586.

Riess, A.G., Filippenko, A.V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P.M., Gilliland, R.L., Hogan, C.J., Jha, S., Kirshner, R.P., Leibundgut, B., Phillips, M.M., Reiss, D., Schmidt, B.P., Schommer, R.A., Smith, R.C., Spyromilio, J., Stubbs, C., Suntzeff, N.B., and Tonry, J. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.*, **116**, 1009–1038.

Rybicki, K.R. and Denis, C. (2001). On the final destiny of the Earth and the Solar System. *Icarus*, **151**, 130–137.

Sackmann, I.J., Boothroyd, A.I., and Kramer, K.E. (1993). Our Sun III: present and future. *Astrophys. J.*, **418**, 457–468.

Spergel, D.N., Verde, L., Peiris, H.V., Komatsu, E., Nolta, M.R., Bennett, C.L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S.S., Page, L., Tucker, G.S., Weiland, J.L., Wollack, E., and Wright, E.L. (2003). First-year Wilkinson microwave anisotropy probe (WMAP) Observations: determination of cosmological parameters. *Astrophys. J. Suppl.*, **148**, 175–194.

Super-Kamiokande Collaboration. (1999). Search for proton decay through $p \to \nu K^+$ in a large water Cherenkov detector. *Phys. Rev. Lett.*, **83**, 1529–1533.

Zeldovich, Ya.B. (1976). A new type of radioactive decay: gravitational annihilation of baryons. *Phys. Lett.*, **A59**, 254.

·3·

Evolution theory and the future of humanity

Christopher Wills

3.1 Introduction

No field of science has cast more light on both the past and the future of our species than evolutionary biology. Recently, the pace of new discoveries about how we have evolved has increased (Culotta and Pennisi, 2005).

It is now clear that we are less unique than we used to think. Genetic and palaeontological evidence is now accumulating that hominids with a high level of intelligence, tool-making ability, and probably communication skills have evolved independently more than once. They evolved in Africa (our own ancestors), in Europe (the ancestors of the Neanderthals) and in Southeast Asia (the remarkable 'hobbits', who may be miniaturized and highly acculturated *Homo erectus*).

It is also becoming clear that the genes that contribute to the characteristics of our species can be found and that the histories of these genes can be understood. Comparisons of entire genomes have shown that genes involved in brain function have evolved more quickly in hominids than in more distantly related primates.

The genetic differences among human groups can now be investigated. Characters that we tend to think of as extremely important markers enabling us to distinguish among different human groups now turn out to be understandable at the genetic level, and their genetic history can be traced. Recently a single allelic difference between Europeans and Africans has been found (Lamason et al., 2005). This functional allelic difference accounts for about a third of the differences in skin pigmentation in these groups. Skin colour differences, in spite of the great importance they have assumed in human societies, are the result of natural selection acting on a small number of genes that are likely to have no effects beyond their influence on skin colour itself.

How do these and other recent findings from fields ranging from palaeontology to molecular biology fit into present-day evolution theory, and what light do they cast on how our species is likely to evolve in the future?

I will introduce this question by examining briefly how evolutionary change takes place. I will then turn to the role of environmental changes that have resulted in evolutionary changes in the past and extrapolate from those past changes to the changes that we can expect in the short-term and long-term future. These changes will be placed in the context of what we currently know about the evolution of our species. I will group these changes into physical changes and changes that stem from alterations of our own intellectual abilities. I will show that the latter have played and will continue to play a large role in our evolution and in the evolution of other animal and plant species with which we interact. Finally, I will turn to a specific examination of the probable course of future evolution of our species and of the other species on which we depend.

3.2 The causes of evolutionary change

Evolutionary changes in populations, of humans and all other organisms, depend on five factors.

The first and perhaps the most essential is mutation. Evolution depends on the fact that genetic material does not replicate precisely, and that errors are inevitably introduced as genes are passed from one generation to the next. In the absence of mutation, evolutionary change would slow and eventually stop.

The effects of mutations are not necessarily correlated with the sizes of the mutational changes themselves. Single changes in the base sequence of DNA can have no effect or profound effects on the phenotype – the allelic differences that affect skin colour, as discussed in Section 3.1, can be traced to a single alteration in a base from G to A, changing one amino acid in the protein from an alanine to a threonine. At the other end of the spectrum, entire doublings of chromosome number, which take place commonly in plants and less often in animals, can disturb development dramatically – human babies who have twice the normal number of chromosomes die soon after birth. But such doubling can sometimes have little effect on the organism.

A fascinating source of mutation-like changes has recently been discovered. Viruses and other pieces of DNA can transfer genes from one animal, plant, or bacterial species to another, a process known as horizontal gene transfer. Such transfers appear to have played little part in our own recent history, but they have been involved in the acquisition of important, new capabilities in the past: the origin of our adaptive immune system is one remarkable example (Agrawal et al., 1998).

The most important mechanism that decides which of these mutational changes are preserved and which are lost is natural selection. We normally think of natural selection as taking place when the environment changes. But environmental change is not essential to evolution. Darwin realized that natural selection is taking place all the time. In each generation, even if the