

TONY CZARNECKI

POSTHUMANS 1

FEDERATE TO SURVIVE!

*WHO WILL LEAD HUMANITY TO DEFEND IT
FROM SUPERINTELLIGENCE, CLIMATE CHANGE
AND OTHER EXISTENTIAL THREATS?*

London, July 2020

POSTHUMANS – VOLUME 1

FEDERATE TO SURVIVE!

©Tony Czarnecki

The right of Tony Czarnecki to be identified as the author of this book has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

*(Central image credit: Probal Roychoudhury,
<https://www.pinterest.co.uk/pin/857091372811588440/>)*

This edition: July 2020

ISBN: 9798654293619

London, July 2020

For any questions or comments please visit:
<http://www.sustensis.co.uk>

To my grandson, Dominik

Table of Contents

TABLE OF CONTENTS	4
FOREWORD TO <i>POSTHUMANS</i> SERIES	7
INTRODUCTION	8
PART 1 EXISTENTIAL RISKS	11
CHAPTER 1 A RAPIDLY CHANGING PACE OF CHANGE	12
WHAT'S SO SPECIAL ABOUT BEING A HUMAN?	12
CIVILIZATION TYPE I.....	15
EXPONENTIAL CHANGE	17
CHAPTER 2 EXISTENTIAL RISKS	21
ASSESSING EXISTENTIAL RISKS.....	21
NATURAL EXISTENTIAL RISKS THREATENING HUMAN RACE	25
UNKNOWN EXISTENTIAL RISKS	27
CHAPTER 3 MAN-MADE EXISTENTIAL RISKS	28
EXTREME CLIMATE CHANGE AND ECOLOGICAL COLLAPSE	28
MITIGATING THE RISK ARISING FROM CLIMATE CHANGE	30
A GLOBAL NUCLEAR WAR	32
MITIGATING A GLOBAL NUCLEAR WAR.....	37
SOFT AND HARD WEAPONIZED AI.....	40
MITIGATING THE RISK OF WEAPONIZED AI OR CYBER SOLDIERS	43
NANOTECHNOLOGY	44
MITIGATING NANOTECHNOLOGY AND EXPERIMENTAL TECHNOLOGY ACCIDENT.....	46
LABORATORY ENGINEERED PANDEMICS AND SYNTHETIC BIOLOGY	46
MITIGATING ENGINEERED PANDEMICS AND SYNTHETIC BIOLOGY	47
CHAPTER 4 HOW CAN RISKS COMBINE INTO EXISTENTIAL THREATS	49
HOW GLOBAL CATASTROPHES CAN BECOME EXISTENTIAL RISKS?.....	49
GLOBAL SOCIAL DISORDER	50
GLOBAL ECONOMIC DISORDER	54
WHAT'S WRONG WITH CAPITALISM?.....	54
MITIGATING ECONOMIC DISORDER RISKS.....	57
GLOBAL POLITICAL DISORDER	58
CHAPTER 5 MAKING THE WORLD A LESS DANGEROUS PLACE	61
THE FINAL DECADE?.....	61
WHAT ARE KEY STRATEGIES FOR MITIGATING EXISTENTIAL RISKS?.....	68
WHAT IS LIKELY TO HAPPEN IF WE JUST KEEP GOING ON?.....	70
PART 2 SUPERINTELLIGENCE	73
CHAPTER 1 FROM INTELLIGENCE TO SUPERINTELLIGENCE	74

DEVELOPING SUPERINTELLIGENCE.....	74
FROM SUPERINTELLIGENCE AND BEYOND	79
CHAPTER 2 THE RISK ARISING FROM A HOSTILE SUPERINTELLIGENCE.....	85
CHAPTER 3 MINIMIZING THE RISKS OF A HOSTILE SUPERINTELLIGENCE.....	90
SETTING THE RULES FOR CONTROLLING SUPERINTELLIGENCE	90
HOW TO TAME SUPERINTELLIGENCE?.....	92
CAPABILITY CONTROL METHOD.....	93
MOTIVATION SELECTION METHOD.....	95
UPLOADING THE VALUES OF HUMANITY TO SUPERINTELLIGENCE.....	98
CHAPTER 4 CAN SUPERINTELLIGENCE HELP REDUCE EXISTENTIAL RISKS? .	100
PART 3 THE NEED FOR THE WORLD GOVERNMENT	103
CHAPTER 1 WHO COULD ACT AS THE FUTURE WORLD GOVERNMENT?.....	104
WHY DO WE NEED THE WORLD GOVERNMENT?	104
SELECTING A CANDIDATE FOR A SUPRANATIONAL ORGANIZATION	109
IS THE EUROPEAN UNION OUR BEST HOPE FOR THE WORLD GOVERNMENT?.....	114
CHAPTER 2 OTHER CANDIDATES FOR THE WORLD GOVERNMENT	117
NATO	117
THE UNITED STATES.....	117
UNITED NATIONS.....	117
CHINA.....	118
RUSSIA.....	120
PART 4 CREATING THE EUROPEAN FEDERATION.....	121
CHAPTER 1 WHAT VALUES FOR THE EUROPEAN FEDERATION?	122
ABOUT VALUES	122
ABOUT RIGHTS.....	123
KEY RIGHTS BASED ON UNIVERSAL HUMAN VALUES.....	124
BALANCING RIGHTS AND RESPONSIBILITIES	125
CHAPTER 2 NO FEDERATION WITHOUT A REFORMED DEMOCRACY	127
THE PRINCIPLES OF DEMOCRACY	127
CONSENSUAL PRESIDENTIAL DEMOCRACY	129
CHAPTER 3 FEDERATE OR DISINTEGRATE	132
THE MISSING WORD IN THE EU VOCABULARY	132
WHY TO HAVE A EUROPEAN FEDERATION?	133
OPTIONS FOR CONVERTING EU INTO THE EUROPEAN FEDERATION	134
THE EUROPEAN FEDERATION AND ITS SUBSIDIARY ZONES	138
ZONE 1 – EUROPEAN FEDERATION CONVERGENCE AREA (EFCA).....	139
ZONE 2 - EUROPEAN FEDERATION SINGLE MARKET (EFSM).....	140
ZONE 3 - EUROPEAN FEDERATION CUSTOMS UNION (EFCU).....	140

Tony Czarnecki *Federate to Survive!*

ZONE 4 - EUROPEAN FEDERATION ASSOCIATION AREA (EFAA) 141

PART 5 TOWARDS A HUMAN FEDERATION..... 143

CHAPTER 1 AN EVER-EXPANDING EUROPEAN FEDERATION..... 144

CHAPTER 2 GLOBAL WELFARE STATE BY 2040? 146

 THE WORLD OF ABUNDANCE IN JUST ONE GENERATION 146

 EDUCATION IN HUMAN FEDERATION..... 147

 THE FUTURE OF WORK 148

 PERSONAL FINANCE 148

 YOUNG FOREVER – REJUVENATION MEDICINE 150

 LIFESTYLE 150

CHAPTER 3 HOW TO FINANCE A GLOBAL WELFARE STATE? 152

 HIGHER THAN PREDICTED GDP GROWTH 153

 RAISING TAXES TO FINANCE BETTER LIFE SATISFACTION 155

 SIGNIFICANT FALL IN PRICES AND A FASTER GROWTH OF REAL INCOME 157

 SUBSTANTIALLY LOWER COST OF GOVERNMENT 160

 WEALTH CAP 161

CHAPTER 4 GLOBAL WEALTH REDISTRIBUTION FUND (GWRF) 162

 BETWEEN A FISH AND A FISHING ROD 162

 STOP ECONOMIC AND CLIMATE CHANGE-ORIGINATED MIGRATION 165

 INSTILLING UNIVERSAL VALUES OF HUMANITY WORLD-WIDE..... 168

 SETTING UP GLOBAL WEALTH REDISTRIBUTION FUND 169

 A MINI WELFARE STATE FOR DEVELOPING COUNTRIES..... 174

CHAPTER 4 THE EUROPEAN FEDERATION BECOMES A HUMAN FEDERATION 177

CONCLUSIONS 179

GLOSSARY 184

BIBLIOGRAPHY 189

FOREWORD TO *POSTHUMANS* SERIES

The *POSTHUMANS* series has its origin in my first book “Who Could Save Humanity from Superintelligence”. That book was addressed primarily to academics, politicians, and readers already familiar with the subjects discussed there. The three books of the series expand those subjects and present them in a way, which may be more suitable for a casual reader.

“**Federate to Survive!**” is about **WHAT** are the existential risks that threaten our **survival**. Superintelligence is the most imminent and supreme existential risk, as is the climate change, both of which have a tipping point in about 2030. Therefore, minimizing those risks should take precedence above any other goals of our civilisation. The secondary theme in this book is **the selection of an organization**, which could start the federalization of the world and guide us through this most perilous period in the humans’ existence. I have used 10 criteria to select the best candidate from the world’s largest countries and organizations, which seems to be the European Union, despite its own problems.

“**Democracy for a Human Federation**” continues from where the previous volume ended, proposing **HOW** we can survive existential threats. We need two elements to achieve that: **Democracy** and a **Human Federation**. The key reason for an urgent deep reform of democracy is to minimize existential risks, including the risk of delivering a malicious Superintelligence, by priming it with new Universal Values of Humanity. This is also an absolute prerequisite for the federalization of the European Union, which is the first step towards **building a Human Federation**. This secondary theme presents three scenarios for the EU’s federalization and an outline of the European Federation’s Constitution.

“**Becoming a Butterfly**” is the third book in the series, asking **WHO** we may become after 2050, assuming we will survive existential threats. Its focus is on **Superintelligence**. This is a mature form of an ever faster and more intelligent, self-learning Artificial Intelligence. If that final product becomes a malicious entity, it may make us extinct in a few decades. However, if we do it right, it will not only protect us from existential risks but also create unimaginable prosperity in the world of peace, and endless possibilities for human self-fulfilment.

Superintelligence will also offer an evolutionary path for humans to become **Posthumans**. The most important in that evolutionary paradigm shift will be a continuous learning and experiencing by Superintelligence until its maturity, about what it means to be a human. At that moment, we will pass the control over our future to Superintelligence. Humans and Transhumans may then gradually merge with Superintelligence as individual digital Posthumans but acting in unison as a single superintelligent being.

INTRODUCTION

There is quite a wide range of subjects covered in this book, the first in the *POSTHUMANS* series. This is unavoidable if the subject is Humanity. Sometimes the subject areas, which I cover, may not be too familiar to you. However, I had to include them, covering them very superficially for the reason of consistency, and to provide as far as possible, a holistic view on who we are as humans. It is a challenging task and I had to make some shortcuts to achieve my overall objective. Should some of the terms used in this book prove a bit challenging to understand, I would recommend browsing the Glossary at the end of the book.

The main theme discussed in this book are **existential risks**. These are the risks, which threaten our survival as a species. Natural existential risks have been occurring regularly throughout the history of hominids. All of them have perished, the last one, the Neanderthals, just over 30,000 years ago. Today, we face similar risks. However, this time these are the existential risks, which we have created ourselves. They may indeed lead to the extinction of a human species.

In the last century, it took about a decade for change in technology to produce a significant societal effect. Today, such change can take just one year and by the end of this decade it will take about one week. This is called **exponential pace of change**, which starts to permeate many domains of our lives. The only pace of change human species has known so far is linear, when change occurs in small, similar steps every year or every month. Therefore, exponential pace of change is almost impossible for most people to absorb and that itself increases the probability of triggering various existential risks. Of those existential risks, Superintelligence is unique, since it can materialize much sooner than, for example, climate change, and which may annihilate all humans.

To understand what **Superintelligence** means, let me first start with **Artificial Intelligence** (AI), about which you may have already heard. It is a digital computer (a machine) imitating human cognitive abilities like problem solving, learning, and understanding of language and speech. Each time you use your car navigation system, Google's Personal Assistant or Amazon's Alexa, you deal with AI, at least to some degree. Such AI can already surpass best human capabilities but in one area only. Therefore, it is usually called a "narrow AI".

Over the next 20-30 years, it will gradually mature into a single intelligent entity, as the most complex, almost invisible network, with all the supporting hardware, and software, surpassing any human being in every skill or task that we could master. It will be able to appear suddenly in several parts of the world

simultaneously, perhaps as a smart gentleman, whom it would be difficult to distinguish from other people walking in the street. Such ability to have any skill and perform any task better than any human is called **Artificial General Intelligence** (AGI). In this book, I use instead the term Superintelligence, best popularised by Nick Bostrom, a well-known authority on AI, in his book under this title.

The need to create a new supranational organization, which would federate the world is another subject that I cover in this book. Nations get together for many reasons. They can stay independent while being part of a confederation, created around a mutual goal, like never fighting each other, so they can all live in peace. The European Union is the best example of that, which has enabled peace in Europe for over 75 years. Nations can also federate around values, in which case, they will by definition live their lives in a similar way by sharing common values, tradition or culture. For that, they are prepared to share part of their nation's sovereignty. The USA, built on the principle of freedom, equality, and social solidarity, is the best example.

Humanity needs to federate as a planetary civilization right now, or the human species may become extinct. However, realistically, it would be impossible to federate all nations based on common values today. This is what the World Federalist Movement has been trying to achieve for the last 75 years, waiting for all nations to live by the same system of human values, such as peace.

In my view, the long-term goal must be to build a global federation and a planetary civilization based on common values. However, if we wanted to do it today, then it would limit the number of the nations that could be quickly federated, since only the nations with a similar culture, and to some extent historical past, may consider joining such a federation. Therefore, any such a federation will initially include only some countries. The foundation of that federation must be a new charter of the Universal Values of Humanity. This is discussed in the book in the context of finding the most suitable existing organization that could initially act as a de facto World Government. I explain in some detail, why the United Nations is, unfortunately, not a suitable candidate.

Only when such an initial federation has been set up, it can then, by creating a critical mass and guided by its own goals, gradually expand its membership, ultimately federating all nations within a Human Federation.

1

PART 1
EXISTENTIAL RISKS

Chapter 1

A rapidly changing pace of change

What's so special about being a human?

It is quite natural that we, humans, think about ourselves as a superior race, the most advanced species that is here to last for ever. Of course, we would like to advance such a conclusion because that in some sense makes us immortal as a group. However, such an assumption is rather unfounded. Think about all those millions of species that have existed on earth and are now extinct (over 99%), including some early human species. Evolutionary biologists using currently available archaeological evidence tell us that all humans evolved from apes over 6 million years ago in Africa. It is believed that there were about 20 different species of early humans. Some of them emerged several million years ago. About 2 million years ago (but some scientists still think it was much later), those humanoids migrated in several waves into Asia, then Europe and later on to the rest of the world. From the most advanced hominids, only *Homo sapiens* has survived. The last group that co-existed with *Homo Sapiens* were the Neanderthals who became extinct about 30,000 years ago⁽¹⁾.

We, humans have plenty in common with other apes and species. In his intellectually stimulating book, "Human Purpose and Transhuman Potential"⁽³⁾ Ted Chiu points out that many differences between us and other mammals are differences of degree. For example, many of our seemingly unique traits are just exaggerated versions of traits that are already identified in other mammals and animals: chimps kiss, laugh, lie, and have in-group politics and show goal-directed action. Ants, wolves, and dolphins all have social traits. Many primates are self-aware. Elephants cry. Capuchin monkeys have forms of monetary exchange, and so on. These are just a few examples, but the point is that many of our behaviours are actually much more advanced versions of innate "animal instincts."

But saying that, Chu maintains that there are three "revolutionary" traits that make us unique and many other biologists and scientists agree with him. These are:

1. **Symbolic abstract thinking:** This is our ability to think about objects, principles, and ideas that are not physically present.
2. **Structure building:** The ability to build physical and social structures, in addition to mental models.

3. **Higher consciousness:** The very fact that humans can write and read articles like this one and contemplate the uniqueness of our mental abilities is awe-inspiring.

There are of course many other ways, in which we differ from animals, depending on the perspective one takes. In “What makes us human? ⁽²⁾”, Lisa Marder selected 10 traits, which I believe could be a good foundation for comparing humans with Superintelligence in the following chapters:

1. **Brain** - This is the most extraordinary feature of humans. Its relative size, scale, and capacity are greater than that of any other species. The human brain’s relative weight to the total weight of a human body is 1:50. Most apes have a ratio of 1:180. That makes the human brain relatively three times bigger than the brain of a gorilla.
2. **Mind** - Most scientists view mind as physically based, emerging from the activity of its neurons and synaptic connections. The human mind is different from the brain, since it most likely consists of consciousness, which definition is not universally agreed but generally understood as a subjective experience of ‘self’ and the world around; the experience of "feeling", thoughts, beliefs and imagination.
3. **The Hand and its Thumbs** - Most primates have opposable thumbs. That means they can be moved around to touch other fingers, which in turn allows grasping things. But the human thumb differs in two ways: in exact thumb’s location and its relative size to other fingers. This has given humans finely tuned motor skills and the ability to engage in detailed precision work necessary to implement creative ideas by technological means.
4. **The Larynx (Voice box)** - This is one of the most crucial differences between us and other mammals. The larynx box is lower in the throat of humans than for example in chimpanzees. Together with the increased flexibility in the mouth, the shape of the tongue and lips it enables us to speak.
5. **The Shoulder** - The human shoulder joint has a wide range of motion, giving us the potential for great leverage and accuracy in throwing objects (important in hunting, which increased the chances of survival).
6. **Standing upright, being bipedal** - This is another important difference that makes humans unique, because it gave us the ability to hold and carry objects, or pick them up and throw, as well as see from a higher vantage point than most apes can do. Once humans stood upright, they were able to cover greater distances while expending relatively less energy.
7. **Naked Hairless Skin** - This may seem an unimportant human trait. However, without it, humans may have not survived major climate change about 200,000 years ago that demanded them to travel long distances for food and water. The evolution created plenty of sweat glands in our skin, which enabled the body to dissipate heat the more efficiently, the less hair covered the skin.

8. **Blushing Response** - This is one of two social traits that make us human. No other mammal has this trait, and it is still unclear what kind of evolutionary advantage it gives us over other species. Some evolutionary biologists, e.g. Frans de Waal, say that "...Blushing interferes with the unscrupulous manipulation of others." (3)
9. **Religion and Awareness of Death** - The second social trait is religion and awareness of death. Humans are unlike other species that live unaware that their existence will end one day. Although some species, e.g. elephants, seem to mourn when one of their family members has died, they themselves are unlikely to know that they will also die.
10. **Biochemical Factors** - Finally, we differ from other animals by our biochemical composition. Scientists have discovered certain biochemical markers and genes that are specific to humans. For example, one such gene, FOXP2, which humans share with the Neanderthals and chimpanzees, is critical for the development of speech and language.

That encyclopaedic definition of Humanity only points to differences with other species within the earthly biological system. When we compare ourselves with the perceived characteristics of the future Superintelligence, which is important in the context of this book, Humanity differs in that it is:

- Anchored to ethics – governed by a system of values
- Culturally motivated – exhibiting its existential sense through culture
- Technologically dependent – enabling its continuous development via scientific and technological progress (there would be some overlap with Superintelligence)
- Expressed itself within several civilizations

Therefore, when we discuss an eventual morphing of humans with Superintelligence, we have to remember our evolutionary traits, which stand out even more when we take a cosmic perspective. We are not only in the universe, but the universe is also within us. Our brains, as an extension of the universe, are now being used to understand themselves. That is why Carl Sagan has famously said "we are a way for the cosmos to know itself." We are the only living being on Earth that can do this. We know these are uniquely human traits.⁽⁴⁾

How will Superintelligence view such a perspective on our role in the Universe and if Superintelligence's perspective could be even more worth pursuing than our, is another question. We may have to answer it in the near future, assuming that we will retain full control over Superintelligence.

Civilization Type I

There are many ways, how we can identify civilizations. If we take an **historic** view, then a differentiating factor for various stages of Humanity's progress would be technology. It is technology that ultimately underpins and differentiates civilizations across the millennia. In broad terms we have had four types of civilizations:

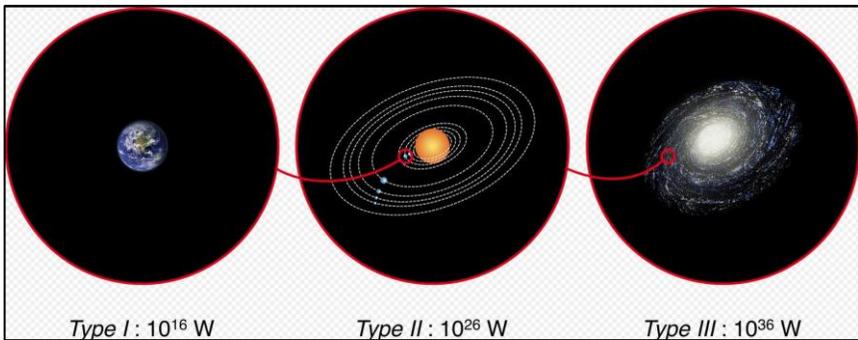
- Tribal/Nomadic
- Agrarian
- Industrial
- Post-industrial (Digital) - today

When viewed from a **geographic** perspective, then at any given time a stronger differentiating factor than technology is culture, customs, and religion. It is true that an earlier use of some technological innovations (e.g. paper in China or print in Europe) may have contributed to important differences between civilisations existing at the same time in different geographical locations. However, it was ultimately the culture and morality stemming from different set of values that shaped those civilisations more than technology. For example, in the Middle Ages there were Aztecs, Mayans, Incas, Chinese, Japanese, Indian and European civilizations – to mention just the most influential ones.

Physicists define civilizations by the **energy** level that could be available for its growth. In 1964 the Russian astrophysicist Nikolai Kardashev defined three such types of civilizations, which differ by the order of energy they had available to them, measured in Watts (W). Each civilization differs from the other by 10 orders of magnitude. Here is a succinct summary of the so-called Kardashev scale (5).

- Type I civilization—also called a **planetary civilization**—can use and store all of the energy which reaches its planet from its parent star.
- A Type II civilization—also called a **stellar civilization**—can harness the total energy of its planet's parent star (the most popular hypothetical concept being the Dyson sphere—a device which would encompass the entire star and transfer its energy to the planet(s)).
- A Type III civilization—also called a **galactic civilization**—can control energy on the scale of its entire host galaxy.

Here is an illustration of how such civilizations might evolve:



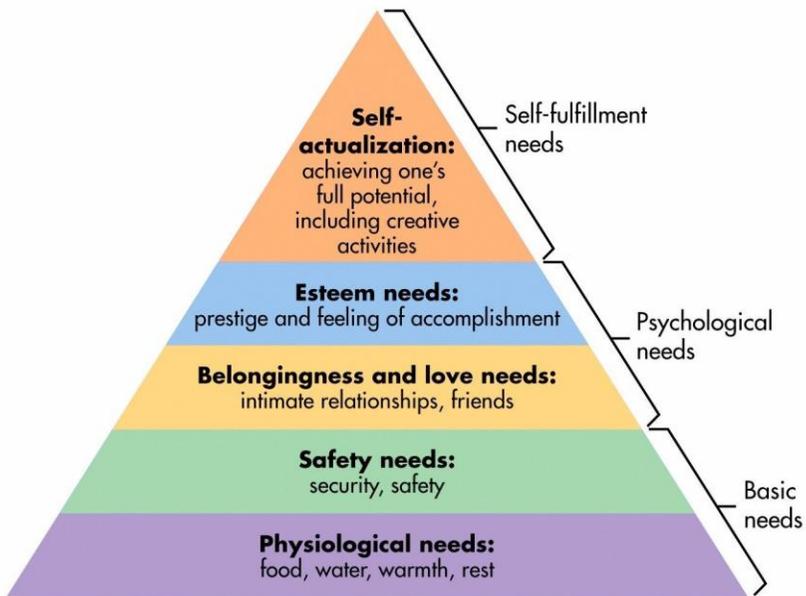
Energy consumption in three types of civilizations defined by Kardashev scale (5)

A similar approach to defining civilization is proposed by Michio Kaku, a renowned US physicist (6). It is a derivative of Kardashev scale but proposes a more generic differentiating factor - the use of **resources**, which of course also includes energy. I quote it here because it is a valuable framework for a potential advancement of Humanity into a new species morphed with Superintelligence or becoming Superintelligence itself.

- **Type 0 Civilization** is essentially our civilization. A type 0 civilization has only just begun to tap planetary resources such as solar power, geothermal power, and wind power. Most of its power generation is still based on non-renewable fossil fuel resources, for example, oil, coal, and natural gases.
- **Type 1 Civilization** can effectively control the entire resources of their planet; they can predict weather patterns and earthquakes very accurately and even control them, by using artificially induced greenhouse effects or space-based lasers. A Type 1 Civilization could conceivably halt an ice-age.
- **Type 2 Civilization** has the capability to extend their power to their entire Solar System by harnessing the power of their suns through Dyson spheres. Having colonized or at least extensively explored all the planets within their Solar System, they are a largely space-faring race and have already mounted expeditions to other stars using interstellar craft.
- **Type 3 Civilization** spans entire galaxies having colonized all the stars by wave after wave of interstellar craft. They can harness the power of galaxies. By utilizing the millions of black holes that are believed to reside within galactic nuclei, type 3 civilizations would have sufficient power to conduct truly universe-changing high-energy physics experiments and examine matter down to the Planck length.

Perhaps the most useful approach to defining civilization and how it could ultimately facilitates human aspirations is the Hierarchy of Human Needs

proposed by Abraham Maslow in his 1943 paper "A Theory of Human Motivation" in *Psychological Review* (7). In his original paper, Maslow's pyramid of human needs starts with a base that identifies an individual's basic **physiological** needs, such as food, sex, or sleep. The next level is **safety and security**, followed by **love and belonging** through to esteem and, finally, at the top of the pyramid to what he called "**self-actualization.**" He further suggested that people who have these needs fulfilled are generally happier than those who don't.



Maslow's Revised Hierarchy of Human Needs (8)

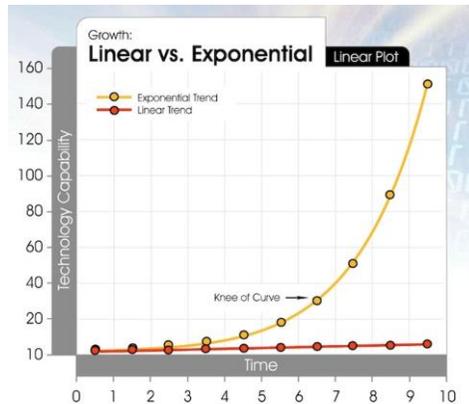
In 2015, a team of psychologists from Arizona State University, including Rick Nauert, tested Maslow's assumptions (9). The researchers found that fulfillment of a diversity of needs, as defined by Maslow, seem to be universal and important to individual happiness. However, the order in which "higher" and "lower" needs are met has little bearing on how much they contribute to life satisfaction and enjoyment.

In summary, the main function of civilizations across the ages has been the creation of better environment and capabilities for the satisfaction of Humanity's needs.

Exponential Change

Change is nothing new, as the Greek philosopher **Heraclitus of Ephesus** said: “Panta rhei” – “everything flows”, or “one cannot enter the same river twice”. So, change is imminently linked to universe and life itself. However, the difference that Humanity has started to experience for the first time in its history is that we have now entered a period where **global change happens at an exponential, rather than a linear pace**. This type of change is called exponential, because at each new moment in time (say every year), the value of what we measure (e.g. speed or growth) would be double that, what it was at the previous moment in time.

To put it more formally, linear change has an exponent of 1, i.e. it multiplies the base number by itself just once. It is called linear because the growth of e.g. number of cars sold will be the same as in the previous period. So, the value of growth is the same in every period. Exponential change with the value of 2, multiplies the base by itself twice.



Source: “Singularity is Near” (10)

Today, the exponential growth of technology is starting to reach the so called “knee of the curve” (see the diagram above). This is the stage at which an exponential trend becomes noticeable. Shortly after this stage the trend can really explode. Therefore, in broad terms:

What now takes a year to complete, in 10 years’ time it will take a week.

Ray Kurzweil precisely makes such an observation saying that we often miss exponential trends in their early stages because the initial pace of exponential growth is deceptive—it begins slow and steady and is hard to differentiate from linear growth. Hence, predictions based on the expectation of an exponential pace can seem improbable. Kurzweil believes that by the middle of the 21st century the exponential curve will be so steep that it will be almost vertical. Such

changes will be so significant that our current level of intelligence is not able to imagine this new paradigm. (11).

To illustrate this point let me quote a fragment of an interview that Seth Teller, one of the top scientists at MIT specialising in driverless cars, gave in 2013 (12):

“Q. Is it possible that in a quarter of a century cars will be self-driven and we will become just passengers?”

A. That is quite a likely scenario, however, I do not think it will be achieved in the next 25 years.”

As most people know, Google self-driven cars have completed over 1m miles in 2015 and in the same year the US Dept. of Transport issued licences for the first driverless lorry. The same happened in the UK in 2016. The Tesla Company and the Volvo Company announced in 2015 they will mass produce fully driverless, so-called 5th level cars, which means completely autonomous vehicles, by the end of 2019. That is 19 years earlier than the prediction given by one of the top specialists in this area. And that is what has happened indeed, although so far only in some cities in the USA e.g. in Phoenix, because of legal restrictions rather than technological failure.

Let me give you a few more examples. Go-Go is a Chinese abstract strategy board game for two players. It is the most challenging game that humans have invented. It has been calculated it has 10^{170} positions – more than the number of atoms in the universe. AlphaGo – a computer program created by Google’s DeepMind Company has for the first time beaten the Grand Master in that game in February 2016. That is a feat in itself. However, more important is that it did it **a decade earlier** than had been predicted than two years earlier.

Then there is the sequencing of the human genome, first completed in 2003 at a cost of about \$3 billion. The next one in that same year costed a little more than \$100M. It’s possible to do it today for less than \$1,000. Perhaps, even more significantly, the cost of sequencing the human genome fell faster than the famous Moore’s Law, which refers to an observation made by Intel co-founder Gordon Moore in 1965. He noticed that the number of transistors per square inch on integrated circuits had doubled every year since their invention (near exponential fall in price). Moore’s law predicts that this trend will continue into the foreseeable future. Since the genome has now been mapped into bits that computers can process, the genome sequencing has in fact become an information technology task. **Human genome sequencing cost now decreases faster than exponentially.** If that trend continues, then the cost of genome sequencing in 2022 may be cheaper than a blood test.

There is another example, reminding us that it may be unwise to claim that a powerful technology will never come to fruition, or is decades away. On

September 11, 1933, Lord Rutherford, one of the most eminent nuclear physicists of his time, described the prospect of extracting energy from atoms as nothing but “moonshine.” Less than 24 hours later, Leo Szilard invented the neutron-induced nuclear chain reaction. Detailed designs for nuclear reactors and nuclear weapons followed a few years later⁽¹³⁾.

Similar examples of much faster than expected discoveries in medicine outside genomics, e.g. cancer treatment, show that even people working at the very forefront of some industries have difficulties seeing that the change is no longer linear.

Exponential change does however have impact not only on individual domains, but also the progress in one domain feeds back to another, such as advances in computer technology impact medicine. This so-called convergence of technologies resulting from the interaction between various parts of individual technologies creates new opportunities that speed up the pace of change much more than a single technology could have done on its own.

So, change is now really happening at an exponential pace. Therefore, wider predictions for what the world might look like say in 20- or 50-years’ time may actually be too conservative. Such exponential change may have either a positive or negative impact on the long-term outcome for the human race. This largely depends on how we use the potential of such discoveries and innovation, i.e. Artificial Intelligence.

Positive changes relate mainly to the unprecedented technological capabilities that could significantly improve the quality of our lives and give a new meaning to what our civilisation is about and how we define the human species. It may also enable the expansion of human race beyond the solar system.

Negative changes, such as global nuclear wars or pandemics may wipe out our civilization in months or lead to a progressive inability to cope with rapid changes that in the course of time will lead to the demise of human species. Therefore, from the human perspective, perhaps the most significant are the changes outside the technological domain, e.g. in social and political domain. For example, China has reduced the number of people in permanent hunger by 600m in just 20 years. Life expectancy increases by 6 hours every day in most advanced societies. These are really great positive changes.

Chapter 2

Existential risks

Assessing existential risks

Over the last decade or so, the world has been experiencing the very first signs of some fundamental shifts in political and social domains, such as the rise of populism that has brought to power some extreme right politicians as well as the first wave of unprecedented number of refugees and migrants from Syria, Iraq, Afghanistan, sub-Saharan Africa etc. This is very visible, and people start asking questions about how changes like this can affect our lives, or indeed the future of Humanity. This can happen when such, initially local and insignificant events, like the waves of migrants trying desperately to cross the Mediterranean Sea, are combined with other risks that would materialize at the same time that may lead to catastrophic global events.

Some catastrophic risks are natural, such as super-volcanos or an impact of a large asteroid that can cause extinction of many species. Earth has experienced such mass extinctions in the past. This also includes global warming, which in the past was natural and today is **anthropogenic i.e. man-made**. Other such anthropogenic risks are pandemics caused by artificially created biologically active viruses, or nuclear wars. Perhaps that's why Nick Bostrom of Future of Humanity Institute believes that human extinction is more likely to result from man-made than natural causes (14).

So, what are the risks that may destroy our civilization and Humanity? In most general terms, these risks could be grouped as global catastrophic or existential (terminal), when classified according to their scope and severity. A "**global catastrophic risk**" is any risk that is "global" in scope, lasting for some time (endurable) and hence may kill the vast majority of life on earth but humanity could still potentially recover. An "**existential**" risk on the other hand, is such that its intensity is terminal, and its effect is transgenerational. It can destroy all human, non-human and even plant life.

An example of a **global catastrophic** risk that could destroy the world in a material sense and potentially eliminate human race is a global nuclear war. It could immediately wipe out most of the current material substance of civilization, i.e. towns, infrastructure, food crops, etc. and in the longer term, through radiation, lack of food and the emergence of post-nuclear winter, causing the death of all remaining people.

A global pandemic is an example of an **existential risk**. It could be caused by an accidental or intentional release from laboratories of a deadly virus that would wipe out all humans in weeks but leave the infrastructure undamaged, at least for a few years.

Not all risks are equal. They differ according to their impact and the probability of the risk materializing, which is calculated using the following formula:

$$\text{Risk} = \text{Probability} * \text{Impact}$$

What is existential risk?

Risk = Probability * Impact



Probability of significant pandemics by 2100 = 0.07%

What is impact?

Examples of Impact Estimation in Overall Risk Calculation					
Examples	Low	Moderate	Significant	Catastrophic	Source
UK Vehicle accidents in 2015	Low - unreported, Skin scratches, pain - est. 1,000,000	Slight injury, e.g. broken arm: 162, 315	E.g. Loss of limbs, sight - 22,144	UK car crash deaths - 1,730	UK Dept. of Transport 2015 Report
Pandemics	HIV from 1981 - 34m - about 0.01% world population	Smallpox in 20th century - 300m out of 11.5 bn living - about 3%	Black death 14th century - 11% world population	Death of all people	Global Challenges Foundation 2016 Report
Impact value	0.25	0.5	0.75	1	

Estimated impact of significant pandemics e.g. Black Death is 0.75%

Thus the overall Significant Risk for human existence caused by natural pandemics by 2100 is 0.75%* 0.07% = 0.0525%

There is no fail proof methodology for calculating such risks, nor is the data as rigid as in the case of a car crash. Therefore, the risks calculated by various organizations or well-known specialists in this area vary significantly.

For example, the Future of Humanity Institute carried out a survey in 2008 of the academics gathered at a conference discussing global existential risks and the likelihood of the most significant anthropogenic risks. They estimated **the overall risk of human extinction this century at 19%**. That means, by the end of this century there is at least 19% chance that one of the existential risks in the table below may materialize with its worst impact scenario, which can potentially be the end of the human species. In the period under consideration in

this book, say over the next generation, such probability is about 5%. These are only the risks over which we have some control, mainly in political, military, and social domains, such as nuclear wars, or Superintelligent agent “runaway”.

(15).

Assessing such risks is very difficult because of the interconnections (sometimes called convergence) between them. Therefore, these numbers have to be really considered as indicative and prone to significant errors, mainly because of lack of hard data or insufficient understanding of the impact and the spread of a specific risk.

Humanity’s Top Existential Risks in 21 st Century		
	Risk	RISK (Probability *Impact) of human extinction by 2100 (%) from an expert survey 2008
	Overall Risk	19%
1	Weaponized AI	5%
2	Superintelligent AI	5%
3	Non-nuclear wars	4%
4	Engineered pandemic and synthetic biology	2%
5	Nuclear wars	1%
6	Nanotechnology accident	0.50%
7	Natural pandemic	0.05%
8	Nuclear terrorism	0.03%

Source: Future of Humanity Institute, 2008 (15)

In Global Challenges Foundation report the risks quoted are very conservative. On the other hand, an influential Stern Review on the Economics of Climate Change (16) calculates the overall risk of human extinction as 0.1% per year. This figure may be considered small but over time it can grow quite significantly, leading to the likelihood of human extinction over a century to about 9.5% (in 2005 when the Review was completed). That means, an average person’s chance of dying in a human extinction event over 100 years is nearly **10 times higher than dying in a car accident**. By the way, Stern Review is explicit that the number isn’t based on empirical considerations but is just a useful assumption. However, even more interestingly, the academics who have considered the evidence, generally offer probability estimates much higher than 9.5%. For example, in his book ‘Our Final Hour’, Sir Martin Rees, a former Astronomer Royal, claims that civilization has a **50/50 chance of making it through the present century** (17). When we compare it against the risk of dying in a car crash, this means that an average person is almost 50 times more likely to see civilization collapse than to die in a car accident. The futurist Nick Bostrom argues that it **“would be misguided” to assign a probability of less than 25%** to an existential catastrophe before 2100, adding that “the best estimate may be considerably higher (18). Thus, all of the estimates mentioned earlier confirm

that we should be more worried about existential risks than any other individual risks, such as dying in a car crash.

Finally, given that future technologies are likely to introduce entirely new existential risks, a discussion of existential risks five decades from now could be dominated by scenarios that are unknowable to contemporary humans, just like nuclear weapons or engineered pandemics were unknowable to people a hundred years ago. This suggests that **even the 50/50 risk of the end of civilization by the end of this century provided by Rees may be an underestimation**, since his figure is based on the analysis of currently known technologies. (19)

Over the last 100 years, the most significant existential threat for Humanity was not the First, or the Second World War, but the 1918 flu pandemic lasting till December 1920. It was an unusually deadly influenza, the first of the two pandemics involving H1N1 influenza virus. It infected 500 million people across the world, including remote Pacific islands and the Arctic. It resulted in deaths of 50 to 100 million (three to five percent of the world's population, making it one of the deadliest natural disasters in human history) and leading to a drop in life expectancy by about 12 years (20).

But that was not the first such pandemic in human history. As the table below shows, some pandemics, such as Black Death might have wiped out up to 17% of the world's population. Smallpox death in the 20th century amounted to about 3% of the population, assuming that the total no. of people born in that century was 11.5bn. If we consider that today we have means of transportation more than a hundred times faster than in 14th century, we should assume that pandemics in the 21st century e.g. similar to Ebola virus, would have been far more deadly.

Historic plagues and pandemics

PANDEMIC	DATE(S)	LETHAL IMPACT
Plague of Justinian	AD 541-542	25-33m (13-17% of the world population)
Black Death	14th Century	50-75m (11-17% of world population)
Smallpox	1520-1527	200,000 deaths within the Aztec population (75% of population in some areas)
Spanish influenza	1918 - 1919	50 - 100m (2.5-5% of population)
Smallpox	20th century	300m over the course of the 20th Century
HIV/AIDS	1981 - present	34m

Source: *Historic plagues and pandemics - Global Challenges Report 2016* (21)

Furthermore, over the last three decades several more existential risks such as nanotechnology or Artificial Intelligence have been added.

Natural existential risks threatening human race

This book is predominantly about anthropogenic (man-made) risks since we can do very little about those risks that do not depend on Human activity (that's why they are called non-anthropogenic risks). However, I shall cover them very briefly for consistency. I will use the risk estimates from Global Foundation Report 2016.

Asteroid impact

An asteroid of at least 5 km across is big enough to end civilisation and wipe out human life. It hits Earth about once every 20 million years. But programs to map hazardous objects are making progress and, given enough warning, a concerted effort by the world's space powers might succeed in deflecting an incoming asteroid on to a non-collision path. According to Global Challenges Foundation, current probability of this risk over this century is 0.00013% ⁽²²⁾

Super volcanos

An eruption ejecting thousands of cubic kilometres of material into the atmosphere, far larger than anything experienced in human history, could lead to a "volcanic winter", with effects similar to an asteroid impact or nuclear war. Such events are known from the geological record to have caused mass extinctions. And with today's technology, there is not much we could do to prevent its effects. Probability: 0.00003% ⁽²²⁾

Cosmic threats

A number of astronomical threats have been identified.

- **A close encounter with massive objects**, e.g. a star, large planet or black hole, could be catastrophic if occurred in the Solar System. ⁽²³⁾
- **A powerful solar flare or solar superstorm**, which is a drastic and unusual decrease or increase in the Sun's power output, could have severe consequences for life on Earth. (24)
- **Geomagnetic reversal**. The magnetic poles of the Earth shifted many times in geologic history. The duration of such a shift is still debated. Theories exist that during such times, the Earth's magnetic field would be substantially weakened, threatening civilization by allowing radiation from the Sun, especially solar wind, solar flares or cosmic radiation, to reach the surface. These theories have been somewhat discredited, as statistical analysis shows no evidence for a correlation between past reversals and past extinctions ⁽²⁴⁾.

If humans disappear because of natural causes such as an extremely large asteroid, then will life become totally extinct as such? Well, it may be the only good news in this chapter because the answer (apparently) is that it would be very difficult indeed to achieve a total life extinction. Rafael Alves Batista and David Sloan, both from Oxford University, tested various conditions that would be necessary to kill the world's hardiest species, the tardigrade, also known as the "water bear" and measuring just 0.5 mm. Tardigrades can survive most difficult conditions such as a brief temperature fall to -272°C (close to an absolute 0C) or heat up to $+150^{\circ}\text{C}$. They can stand atmospheric pressure to over 1,000 times the pressure at sea level. They can survive in a complete vacuum for up to 30 years without food or water. They can also survive radiation doses thousands of times the doses that would have killed humans. So, the question those two researchers tried to answer was what type of cataclysmic events might be able to finally kill off the hardy tardigrade, i.e. what would need to happen to destroy every living thing on the planet?

Here is their answer:

"...the planet's entire oceans would have to boil. On Earth, this would require an incredible amount of energy – 5.6×10^{26} joules (around a million years of total human energy production at current rates). Therefore, we have to consider the astrophysical events, which could provide such an enormous amount of energy" ⁽²⁵⁾. They selected three primary candidates that could do this:

- asteroid impacts,
- supernovae and
- Gamma-ray bursts.

To be hit by an asteroid that would kill literally all life on Earth, is very unlikely, since it might happen once in 10^{17} years, i.e. longer than the age of the Universe. Considering the rates at which supernovae occur, sterilisation is unlikely to happen more than once in 10^{15} years, i.e. far beyond the age of the universe.

Finally, gamma-ray burst is another candidate to eliminate life on Earth completely. These are explosions producing enormous amounts of energy focused into jets of radiation as narrow as a couple of degrees. Analysing these bursts the authors found that they could only kill off life on an Earth-like planet if their origin was within 42 light-years and the planet lay within the beam and there are no candidates for such a gigantic source of energy within that range ⁽²⁵⁾.

Therefore, even if some risks listed in this section do occur, leading to the extinction of humans, life on Earth is likely to continue as it has done for over 3.5 billion years.

Unknown existential risks

This is a catch-all category to cover the unknown unknowns, such as nuclear terrorism. As unprecedentedly powerful technologies are becoming more accessible, the global population is growing, meaning that the absolute number of malicious agents could increase proportionally. According to an American psychologist Martha Stout, roughly 4 percent of the global population are sociopaths. This translates to about 296 million sociopaths today, and if the population rises to 9.3 billion by 2050, this number will increase to 372 million. Although not all sociopaths are violent, they are disproportionately represented among groups such as prison inmates and dictators. It follows that this demographic could seriously jeopardize our collective future if nuclear weapons, biotechnology, nanotechnology, or some as-yet-unknown technology were to fall into the wrong hands.⁽²⁶⁾

The menace posed by ideological extremism is also growing. For example, the number of hate groups in the United States rose from 457 to 892 between 1999 and 2015. Outside the United States, the number of Salafi-jihadist organizations rose from 3 in 1988 to 49 in 2013, the year before the Islamic State emerged as arguably the largest terrorist organization in human history. There are strong reasons for expecting the total population of radical extremists of all political and religious persuasions to continue increasing, due in part to the conflict-multiplying effects of global catastrophes like climate change and biodiversity loss. If empowered by advanced technologies, anyone of these individuals, or groups could wreak unprecedented havoc on a society⁽²⁷⁾.

Nicholas Georgescu-Roegen, a progenitor in economics and a paradigm founder of ecological economics, has argued that the carrying capacity of Earth, i.e. Earth's capacity to sustain human populations and consumption levels, is bound to decrease sometime in the future as Earth's finite stock of mineral resources is presently being extracted and put to use. Consequently, the world's economy as a whole is heading towards an inevitable future collapse, leading to the demise of human civilisation itself.⁽²⁸⁾

In my view, this is quite an unlikely scenario if we consider that Earth is no longer limited by the law of entropy. Earth's entropic limitation would only be true if we could not migrate outside the planet. However, within 50 years there might be 1M people living on Mars (courtesy of Mr. Elon Musk). Such trend would counterbalance this catastrophic prediction.

Chapter 3

Man-made existential risks

Extreme climate change and ecological collapse

This is probably the most publicised existential risk, apart from a global nuclear war. Conventional modelling of climate change induced by human activity (adding carbon dioxide to the atmosphere) has focused on the most likely outcome: global warming by up to 4C. But there is a risk that feedback loops, such as the release of methane from Arctic permafrost, could produce an increase in temperature of about 6C or more. Mass deaths through starvation and social unrest could then lead to the collapse of civilisation. The most optimistic predictions estimate that the overall existential risk from extreme climate change is about 0.01% annually, which would make it 1% over the entire century – not that much. The most realistic assessment was probably made in the Stern Report. It estimates such risk at 9.5% over this century. I have taken the median view, that the risk stemming from the extreme climate change over this century is about 5%.

Martin Rees, the former Royal Astronomer, observes that many people still hope that we can sail towards a low-carbon future without trauma and disaster. He says that politicians won't gain much resonance by advocating a bare-bones approach that entails unwelcome lifestyle changes – especially if the benefits are far away and decades into the future. There are however three politically realistic measures that should be pursued. First, all countries could promote measures that actually save money – better energy-efficiency, better insulation of buildings and so forth. Second, efforts could focus on the reduction of pollutants, methane, and black carbon. These are minor contributors to global warming, but their reduction would (unlike that of CO₂) have more manifested local side-benefits – especially in Asia. And third, there should be a step change in research into clean energy; why shouldn't it be on a scale comparable to medical research? (29)

Martin Rees believes, there is little risk of a catastrophe within, say 50-year time-horizon, so unsurprisingly many politicians downplay the priority of addressing climate change. But if you extend the horizon into the 22nd century and beyond – then you may deem it worth making an investment now, to protect those future generations against the worst-case scenario and to prevent triggering really long-term changes like the melting of Greenland's ice.

He also believes that “political efforts to decarbonise energy production won't gain traction and that the CO₂ concentration in the atmosphere will rise at an

accelerating rate throughout the next 20 years”. By then we will know with far more confidence – perhaps from advanced computer modelling but also from how much global temperatures have actually risen, how strongly the feedback from water vapour and clouds amplifies the effect of CO₂ itself in creating a ‘greenhouse effect’. If the effect is strong and the world’s climate consequently seems on a trajectory into dangerous territory, there may then be a pressure for ‘panic measures. These would have to involve a ‘plan B’ – being fatalistic about continuing dependence on fossil fuels but combatting its effects by some form of geoengineering (Martin Rees ‘The world in 2050 and beyond’)⁽²⁹⁾.

One of the consequences of extreme climate change could be a full collapse of the global ecosystem, so that the planet could no longer sustain a population of billions. This seems to be one of the most complex risks to assess. An ecological disaster, such as world crop failure and collapse of ecosystem services, could be induced by the present trends of overpopulation, economic development, and non-sustainable agriculture⁽³⁰⁾.

Most of these scenarios involve one or more of the following: an extinction event, scarcity of water that could lead to approximately one half of the Earth’s population being without safe drinking water, pollinator decline, overfishing, massive deforestation, desertification, or massive water pollution episodes. A very recent threat in this direction is a bee colony collapse disorder, a phenomenon that might foreshadow the imminent extinction of the Western honeybee. As the bee plays a vital role in pollination, its extinction would severely disrupt the food chain.⁽³¹⁾

There is plenty of coverage of the risks that are linked to climate change. I would not in any sense like to downplay that risk, since it is really multifaceted and not just limited to temperature rise, although this is the major source of the consequences of climate change. However, as I have already mentioned, by the time climate change might really endanger human species and most other species on our planet, which is in the next century, our civilisation will either survive or will most probably be gone because of other risks. Therefore, we should put all our efforts to minimise the risks stemming from Superintelligence because if we make it benign and friendly, it will be our major hope for reducing or entirely eliminating other anthropogenic existential risks. It is probably too late for withdrawing from carbon-based economy. The world will not come together sufficiently quickly and with substantially deep reforms. But not all is lost. In a few decades we will have Superintelligence that will help us deal with this problem and many more. The problem is not how to survive the climate change by the end of this century but how we can survive at all in the next 20 years.

Many climatologists believe that realistically, we must halt the average global temperature increase at 1.5C, and we are already quite close to that threshold. If

we want to achieve the goal of a maximum global temperature of less than 1.5C then we must start dropping CO2 emissions well before 2030. If we continue emitting fossil fuels as before, continuously increasing CO2 level, then by 2030 the Earth's climate may reach a tipping point in many areas, leading to an uncontrolled global temperature increase.

Mitigating the risk arising from Climate Change

Climate Change and the arrival of Superintelligence are two existential risks that will increase progressively over decades rather than emerge within days. Climate Change risk mitigation strategies has been widely covered for over 30 years and therefore I will only make a reference to the most important agreement made recently on mitigating that risk. At the Paris climate conference (COP21) in December 2015, 195 countries adopted the first-ever universal, legally binding global climate deal. It sets out a global action plan to put the world on track to avoid dangerous climate change by limiting global warming to well below 2°C. Key elements of the Paris Agreement cover six areas as summarized by EC Climate Change ⁽³²⁾. The Governments agreed the following actions:

Reduce Emissions by:

- setting a long-term goal of keeping the increase in global average temperature to well below 2°C above pre-industrial levels;
- aiming to limit the increase to 1.5°C, since this would significantly reduce risks and the impacts of climate change;
- agreeing on the need for global emissions to peak as soon as possible, recognising that this will take longer for developing countries;
- undertaking rapid reductions thereafter in accordance with the best available science.

Before and during the Paris conference, countries submitted comprehensive national climate action plans. These are not yet enough to keep global warming below 2°C, but the agreement traces the way to achieving this target.

Ensure Transparency and Global Stocktake by:

- coming together every 5 years to set more ambitious targets as required by science;
- reporting to each other and the public on how well they are doing to implement their targets;
- tracking progress towards the long-term goal through a robust transparency and accountability system.

Promote Adaptation by:

- strengthening societies' ability to deal with the impacts of climate change;
- providing continued and enhanced international support for adaptation to developing countries.

Minimize Loss and Damage by:

- recognising the importance of averting, minimising, and addressing loss and damage associated with the adverse effects of climate change;
- acknowledging the need to cooperate and enhance the understanding, action and support in different areas such as early warning systems, emergency preparedness and risk insurance.

Recognize the Role of cities, regions, and local authorities by:

- scaling up their efforts and support actions to reduce emissions;
- building resilience and decreasing vulnerability to the adverse effects of climate change;
- upholding and promoting regional and international cooperation.

Support non-Party stakeholders

The agreement recognises the role of **non-Party stakeholders** in addressing climate change, including cities, subnational authorities, civil society, and the private sector:

- The EU and other developed countries will continue to support climate action to reduce emissions and build resilience to climate change impact in developing countries.
- Other countries are encouraged to provide or continue to provide such support voluntarily.
- Developed countries intend to continue their existing collective goal to mobilise USD 100 billion per year by 2020 and extend this until 2025. A new and higher goal will be set for after this period.

There are at least 100 various geoengineering techniques that would not damage the environment. One of the best examples is “an artificial tree” developed by Dr Klaus Lackner from Columbia University. In an interview with Yale Environment 360, he says that:

“With \$10 million to \$20 million worth of engineering R&D, we can get off the ground. My hope would be that we then would have a device that can take out a ton a day of carbon from the atmosphere. If you take out a ton a day, you will

need 100 million air capture devices to take out all the CO₂ that we are putting into the atmosphere today. And I would argue that it would be a lot less than that because we would also be capturing carbon at the flue stack, and not making the CO₂ in the first place by developing solar and wind technologies” (33). At this stage, Lackner’s device would cost about \$30,000. If we need 100 million devices, the total cost would be about \$30 trillion. That’s a lot – about 40% of the world’s annual GDP in 2016. So, the only problem of such a solution is their gigantic cost.

However, there are cheaper solutions. For example, in 2016 Harvard prof. Daniel G. Nocera developed what he calls “Bionic Leaf”, a genetically engineered bacterium, which converts sunlight ten times more efficiently than plants. At the same time, it could absorb CO₂ at a fraction of that cost and produce hydrogen⁽³⁴⁾.

Martin Rees proposes other solutions, which may not be that clean but much cheaper and acting very quickly. The ‘greenhouse warming’ could be counteracted by (for instance) putting reflecting aerosols in the upper atmosphere or even vast sunshades in space. It seems feasible to throw enough material into the stratosphere to change the world’s climate. Indeed, what is scary is that this might be within the resources of a single nation, or perhaps even a single corporation. The political problems of such geoengineering may be overwhelming. There could be unintended side effects. Moreover, the warming would return with a vengeance if the countermeasures were ever discontinued; and other consequences of rising CO₂ (especially the deleterious effects of ocean acidification) would be unchecked. So, geoengineering would be an utter political nightmare: not all nations would want to adjust the thermostat the same way. Very elaborate climatic modelling would be needed in order to calculate the regional impacts of any artificial intervention.⁽²⁹⁾

A Global Nuclear War

A nuclear war between the US and Russia was the chief apocalyptic fear of the late 20th century. That threat may have reduced but, with proliferation of nuclear weapons, there is still a risk of a conflict serious enough to cause a “nuclear winter” as the smoke in the stratosphere shuts out sunlight for months. That could put an end to civilised life regardless of the bombs’ material impact. Therefore, it is so difficult to assess the probability of global nuclear war ever taking place and even more difficult to tell if it will ultimately lead to a total collapse of civilisation. That’s why Global Challenges Foundation Report 2017 puts the risk between 1 and 9.5%⁽³⁵⁾, which I have averaged to 5% and this is the level I would consider in further assessment of that risk.

The scenarios that have been explored most frequently are nuclear warfare and doomsday devices. Although the probability of a nuclear war per year is slim, Professor Martin Hellman described it as inevitable in the long run. Inevitably there will come a day when civilization's luck runs out (36). During the Cuban missile crisis, U.S. President John F. Kennedy estimated the odds of nuclear war as being "somewhere between one out of three and even" (37). To put it in today's context, the United States and Russia have a combined arsenal of 14,700 nuclear weapons, and there is an estimated total of 15,700 nuclear weapons in existence worldwide (38).

While popular perception sometimes takes nuclear war as "the end of the world", experts assign low probability to human extinction from nuclear war (39). In 1982, Brian Martin estimated that a US-Soviet nuclear exchange might kill 400-450 million people directly, mostly in the United States, Europe and Russia and maybe several hundred million more through follow-up consequences in those same areas (40). Nuclear war could yield unprecedented human death tolls and habitat destruction. Detonating such a large amount of nuclear weaponry would have a long-term effect on the climate, causing cold weather and reduced sunlight that may generate significant upheaval in advanced civilizations (41).

The most recent scenarios give estimates on the climate change caused by a global nuclear war, without giving the number of casualties. One of these predicts that the explosion of 1,800 US - Russian warhead would cause a long-lasting cold period with a peak average global cooling of 4°C, whilst a larger scale nuclear war with over 3000 warheads (which is only about 35% of the current Russian nuclear stockpile – TC) would cause average cooling of 8°C. This is greater than the average cooling of 5°C experienced during the last ice age, so this would be a severe nuclear winter lasting a decade. Whilst an average cooling of only a few degrees may not sound very serious, the crucial impact is much longer periods of frost in winter and severe drought. There would be dramatically reduced growing seasons or even the impossibility of growing any crop as planned. Farming also relies upon supplies of fuel for mechanised planting and harvesting (42).

However, use of a nuclear weapon today would be much worse for two reasons:

1. A typical modern nuclear weapon is now 8 to 80 times larger; modern society is much more reliant on vulnerable information technology and long-distance supply routes for food and fuel.
2. Modern society is heavily reliant on electricity to power central heating pumps, to provide water, information via TV, the internet, and mobile phones. Nuclear strike will mean no water supply, no heating or lighting, no information, no mobile phone signal.

3. Only a few days of food supply exists in regional distribution depots. The supply network would fail for multiple reasons: road blockages, communications breakdown, collapse of the banking system, destruction of ports.

International aid organisations and health bodies all agree that the tens of thousands of casualties from just one nuclear bomb would overwhelm all attempts to help the injured. As a result, there would be no hope of treatment for severe injuries including burns, broken bones, and deep cuts from flying debris. With the intense levels of damage, huge fires would spread across all major towns and other targets lasting days to weeks. We now understand that these huge fires would cause long lasting climatic impacts at a global level, creating a nuclear winter. Realistically, after a large-scale nuclear war, one must picture small groups of brutalised, traumatised people, violently thrown back into a pre-industrial age. Assuming that some people somewhere furthest from the bombs could initially survive this global catastrophe, any 'recovery' would surely be measured in hundreds of years. It has to be regarded as a shocking indictment of our modern civilisation that current stockpiles of nuclear weapons are sufficient to cause such a global catastrophe^{e (42)}.

Overall, the greatest possibility of a nuclear war today is in Asia. First, it is a significant potential for a nuclear conflict between India and Pakistan. At the heart of this conflict is, of course, the territorial dispute over the northern Indian state of Jammu and Kashmir, which Pakistan says should be its territory. The reason behind a high risk of nuclear conflict between these two countries is the fact that India's conventional capabilities are vastly superior to Pakistan's. Consequently, Islamabad has adopted a nuclear doctrine of using tactical nuclear weapons against Indian forces to offset the latter's conventional superiority. The latest climate models predict that the use of a just few tens to a hundred of the smaller nuclear weapons in the regional India-Pakistan scenario would cause severe frosts, reduced growing seasons, drought and famine lasting up to ten years across the entire northern hemisphere (43). This situation is similar to that between the U.S.-led NATO forces and the Soviet Union during the Cold War. Numerically, the Soviet army was superior to that of NATO. Therefore, the United States, starting with the Eisenhower administration, turned to nuclear weapons to defend Western Europe from a Soviet attack

Another area of a potential conflict is between Israel and Iran. Although Iran apparently does not have nuclear weapons yet, but only some facilities, like plutonium generating centrifuges, it is highly likely, it can build them pretty soon despite the current oversight of the IAEA. Israel has at least once in the past try to stop the Iranian nuclear programme. In June 2010 an advanced computer worm called Stuxnet, was discovered, which is estimated that it might have

damaged as many as 1,000 centrifuges (10% of all installed) in the Natanz enrichment plant⁽⁴⁴⁾.

Israel believes that Iran still continues to develop its nuclear programme deep in the mountains. Western defence experts point to the Iranian Fordo facility, which is located deep underground near the city of Qom, as a site that was immune to conventional air strikes. That is why Israeli leaders have concluded that conventional air strikes would be insufficient in curbing Iran's nuclear program, leaving only a deployment of tactical nuclear weapons or ground forces⁽⁴⁵⁾.

Since summer 2017, the world's attention has been firmly fixated on North Korea. On 4th July, 2017 North Korea launched its first intercontinental ballistic missile which could reach the mainland United States. That has led a few weeks later to a decision by the Hawaiian authorities announcing that they would revive a network of Cold War-era sirens, to alert the public in the event of a nuclear strike. Then on 3rd September 2017 North Korea tested a nuclear weapon far larger than any it had used. James Mattis, the U.S. Secretary of Defence, said it was seven times the size of the bombs dropped on Hiroshima and Nagasaki.

That incident on its own may not have been that scary, had the potential damage only related to a direct impact zone (depending on the bomb size from a diameter of 20 km to 60-100 km). It would require the bomb to actually hit the ground. This seems to be a formidable barrier because on the re-entry the rocket's head might heat up to about 2000C and explode before hitting the target. That's why complicated thermal shields are needed, which N. Korea apparently does not have.

However, in October 2017, two members of the disbanded US congressional Electromagnetic Pulse (EMP) commission said at the House Homeland Security subcommittee hearing that "a nuclear EMP attack from Kim Jong Un was the biggest threat to the US yet". They added: "it could shut down the US electric power grid for an indefinite period, leading to the death within a year of up to 90 per cent of all Americans... A nuclear EMP attack of just a few bombs to cover the whole USA would be needed, to completely make out of use power grids and other critical infrastructures that make modern civilization, and life itself, possible. Eventually, millions would die from starvation, disease, and societal collapse"⁽⁴⁶⁾.

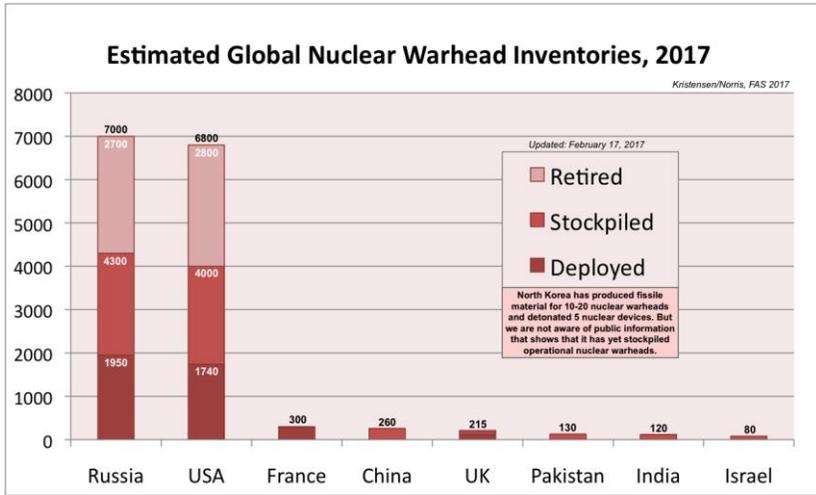
Those who think that North Korea's missiles and nuclear facilities can be wiped out in a preventive air strike would make a potentially disastrous error. N. Korea's nuclear arsenal is dispersed and hidden throughout the country's mountainous terrain. Failing to hit them all would leave some 10 million people in Seoul, 38 million people in the Tokyo vicinity and tens of thousands of US military personnel in northeast Asia vulnerable to missile attacks with either conventional or nuclear warheads. Even if the US managed to wipe out

everything, Seoul would still be vulnerable to attacks from North Korea's artillery ⁽⁴⁷⁾.

The EMP effect of nuclear bombs has been known since the late 1950'. Today, it is almost certain that every country that has the capability to launch nuclear bombs, considers options to use EMP in several ways. The most 'benign' fact is that the number of immediate casualties could be relatively small if EMP bomb is exploded in the near space (above 100-150 km). The casualties would result mainly from airplane catastrophes, car accidents or hospitals. This is the most likely scenario if North Korea uses such a bomb, destroying not only electronic-dependent infrastructure in the USA but also in the neighbouring South Korea. The EMP impact would be comparable with the so-called neutron bomb built under President Reagan administration in 1981. Its main destructive power comes from radiation rather than from an electromagnetic pulse. It would have a similar effect, i.e. causing not too many direct human casualties but an almost complete immobilisation of the advancing army. In the 1980' it was the Soviet army that would have been disabled within a few kilometres of the neutron bomb's impact.

The second option might be to use EMP-type nuclear device in conjunction with other more 'conventional' weapons by one of the nuclear power states such as China, Russia, the USA, France, or the UK. That is however, unlikely because it could have sparked of an all-out 'hot' nuclear global war, since each of these countries has part of its nuclear arsenal dispersed on the always-in the-air airplanes and on the submarines, which could deliver a retaliatory attack. That's the same principles of mutual self-destruction that kept the world nuclear war free during the Cold War period.

The third option is launching such a bomb, with an objective to eliminate advanced military capabilities (e.g. nuclear sites, cyber spying equipment, etc.). That is quite likely in my view by one of these countries: Israel, Iran, India, and Pakistan. That's why the risk of a local nuclear incident is rated by experts as 50:50 in the next decade. A good example is Bill Perry who was a nuclear weapons expert serving as President Clinton's Secretary of Defence. In April 2017 he said that the worst-case scenario would now mean nothing short of a total Armageddon. "An all-out general nuclear war between the United States and Russia would mean no less than the end of civilisation". He believes that today's global security threats are very different to what they once were. The most likely scenario for a nuclear attack would be if a terrorist group got hold of a small amount of enriched uranium, allowing them to make an improvised nuclear bomb. He estimates the chance of this happening within the next 10 years as even ⁽⁴⁸⁾. Yes, the world still has plenty of nuclear weapons that could annihilate all life on Earth as this summary illustrates:



Source: FactCheck Q&A: *Could there be a nuclear Armageddon?* (48)

All the potential nuclear wars mentioned so far are deemed “local”. However, this is highly unlikely if the protégé of some of the ‘big party’ like China were in ultimate danger, in this case North Korea. The same may be said about Israel and Iran conflict, where the probability of serious military engagement, including the use of nuclear weapons, by Russia – supporting Iran, and America – supporting Israel, must be considered. Additionally, one has to include Russia’s current unpredictability in its territorial aims in Eastern and Central Europe. Like in the Cold War time, the Russians have superior manpower and also outnumber NATO in tanks. That might force NATO to use initially tactical, and if a local war gets out of control, strategic nuclear weapons starting a global nuclear conflict.

Mitigating a global nuclear war

There have certainly been efforts for the last 40 years to reduce the risk of a nuclear war. For example, since 2010 every two years Nuclear Security Summits are held. Russia participated in the first three, but withdrew from the fourth one in 2016, blaming the USA for ‘taking over the leadership’. Probably the most comprehensive Summit was the one held in 2014 in the Netherlands, which was attended by 58 world leaders (5 of which from observing international organizations). The representatives attending the summit included U.S. President Barack Obama and Chinese President Xi Jinping. One concrete result of that summit was the progress reached in eliminating stocks of Highly Enriched Uranium (HER). It was reported at the Summit that between 2012 and 2014 15 metric tons of HER had been down blended to Low Enriched Uranium (LEU),

which will be used as fuel for nuclear power plants, equivalent to approximately 500 nuclear weapons.

So, there have been undoubtedly efforts to reduce the risk of nuclear war, but they are not comprehensive enough and are introduced too slowly. There have been quite a few proposals on how to reduce the risk of nuclear wars. So, it is not the lack of knowledge that lies at the bottom of slow progress in risk reduction. It is the usual procrastination and mistrust that rather decreasing, actually increases the risk of a nuclear war. The best example are the five proposals put forward in 2013 by the two former US Secretaries of State, Henry Kissinger and George P. Shultz, and two US National Security Advisers, Sam Nunn and William J. Perry for reducing the risk of nuclear wars, mainly addressing the US-Russia relationships:

1. **Securing nuclear materials to prevent catastrophic nuclear terrorism.** Materials necessary for building a nuclear bomb today are stored at hundreds of sites in 28 countries—down from over 40 countries just 10 years ago. But many of these sites aren't well secured, leaving the materials vulnerable to theft or sale on the black market. Important commitments were undertaken to secure nuclear materials and improve cooperation during the 2010 and 2012 Nuclear Security Summits. **Yet no global system is in place for tracking, accounting for, managing and securing all weapons-usable nuclear materials.**
2. **Changes in the deployment patterns of the two largest nuclear powers** to increase decision time for leaders.
3. **Actions following New START.** The progress in the strategic field has been considerable. But there should be further progress made under the New START agreement with the following prerequisites:
 - Strict reciprocity
 - Demonstrable verification
 - Providing adequate and stable funding for the long-term investments required to maintain high confidence in stored nuclear arsenal.
 - **Consolidating and reducing U.S. and Russian tactical nuclear weapons** not covered under New Start should be a high priority. The nuclear programs of North Korea and Iran undermine the Non-Proliferation Treaty and pose a direct threat to regional and global stability. Unless these two states are brought into compliance with their international obligations, their continued nuclear programs will erode support for non-proliferation and further nuclear reductions.
4. **Verification and transparency of nuclear-security** agreements is absolutely essential to have confidence that they work properly. The U.S.

should launch a "verification initiative" that involves the U.S. nuclear weapons laboratories and global scientific experts in developing essential technologies and innovations for reducing and controlling nuclear weapons and materials.⁽⁴⁹⁾

The second proposal of the former US Secretaries of State, was referring to a detailed plan put forward by the Union of Concerned Scientists, engaged for years in developing proposals for reducing the risk of accidental nuclear wars. It points out that despite the Cold War ending decades ago, the United States and Russia still keep hundreds of nuclear weapons on high alert, ready to launch. This rapid launch option, called 'hair-trigger alert', significantly raises the risk of three types of unintended nuclear attack:

- An accidental launch could occur through a system error
- An unauthorized launch is a deliberate launch that would take place without a presidential order, perpetrated by either insiders or outsiders (e.g. through a cyberattack) or mistaken nuclear attack
- A mistaken launch would be authorized by the president, but in response to a false warning of an incoming attack.

Because land-based nuclear missiles are stored at known locations, they are vulnerable to attack. Hair-trigger alert was originally intended to ensure that, were the missiles targeted, they could be launched in retaliation before being destroyed. Yet the United States' retaliatory capabilities are already ensured by hundreds of nuclear weapons stored on submarines, which can't be targeted. As a deterrent, the high alert status of U.S. land-based missiles is therefore irrelevant. Unlike land-based missiles, nuclear-armed submarines can't be targeted while at sea, providing a credible nuclear deterrent.

Therefore, the nuclear powers should de-alert land-based nuclear missiles. For example, in case of the U.S., it should eliminate options for quickly launching missiles on warning of attack and take its missiles off hair-trigger alert. At present, land-based intercontinental ballistic missiles (ICBMs) in the United States are stored in underground siloes and controlled by nearby launch centres. Removing them from hair-trigger alert could be as simple as manually activating a safety switch that prevents the missile from being launched (these switches already exist and are used by maintenance crews). When the next false alarm occurs, leaders would then not be under the same pressure to launch, eliminating the risk of a mistaken launch, and reducing the risk of an accidental or unauthorized launch⁽⁵⁰⁾.

The continuing risk posed by nuclear weapons remains an overarching strategic problem, but the pace of work doesn't now match the urgency of the threat. The

consequences of inaction are potentially catastrophic, and nations must continue to move beyond their particular interests for the benefit of the whole Humanity. The most recent developments like those in North Korea and on the Indian subcontinent, not to mention the increased risk of cyberattack on nuclear control centres, make nuclear war far more probably than at any time since the end of the Cold War.

Soft and Hard Weaponized AI

I could not find a complete definition of a weaponized AI because such a new subject means many things to many people. However, for the purpose of our discussion I would define **Weaponized AI as any evil, malicious or destructive action addressed either at an individual, a group of people or a country using the AI technology**. There are two approaches: **Soft Weaponized AI**, which uses software applications that achieve malicious objectives by usually compromising or blackmailing individuals through publication of documents, pictures or breaking into security systems, and **Hard Weaponized AI** that directs specialized weapons or equipment at a pre-planned target.

Soft weaponized AI

Let me start with the Soft Weaponized AI. Here, probably the best and most succinct list of what soft weaponized AI could do has been created by a well-known futurist Thomas Frey, who presents a simple scenario: “Virtually every situation presents an opportunity for a weaponized A.I., but each will require different strategies, targets, and techniques. Once a clear objective is put into place, the A.I. will use a series of trial and error processes to find the optimal strategy. A.I. tools will include incentives, pressures, threats, intimidation, accusations, theft, and blackmail. All can be applied in some fashion to targeted individuals as well as to those close to them.”⁽⁵¹⁾ He himself had doubts whether he should publish his list of 36 examples, of what it could involve, because it might give hints to perpetrators. But he concluded that anything he could think of, terrorists and evil doers could come up with such a list as well.

Organization-wide or country-wide Soft Weaponized AI Scenarios. These have already been ‘tested’ at the lowest possible level of threat, using so-called fake news and very primitive AI support, in the elections in the USA and other countries. Now imagine applying more sophisticated AI such as the latest version of Deep Mind’s AlphaGo Zero.

1. **Hijacking a City.** Every city is made up of interdependent systems that function symbiotically with their constituency. Stoplights, water, electric, sewage, traffic control, garbage removal, tax assessment, tax collection,

- police, and fire departments are just a few of the obvious trigger points. Once A.I. can disable a single city, it can easily be replicated to affect many more.
2. **Destroying a Country.** At the core of every country are its financial systems. Weaponized A.I. could be directed to attack essential communication and power systems. Once those are disabled, the next wave of attacks could be focused on airports, banks, hospitals, grocery stores, and emergency services. Every system has its weakest link and this kind of exploitive weaponry could be relentless.

Intimidating Professionals

In any society there are “people of influence” who are critical for maintaining the systems, business operations, and processes that govern our lives. These individuals become the most “at risk” for becoming a target of weaponized A.I.:

1. Stock Analysts—The value of our entire stock market hinges on the assessment of a few key individuals
2. Politicians – Any elected official can be bullied into voting in favour of a specific bill or funding proposal
3. Judges - The outcome of most court cases is decided by a single judge
4. Newspaper Editors - These people decide what goes on the front page
5. Corporate CEOs – The CEOs are a huge factor in determining the success or failure of a business
6. Medical Doctors – Doctors and physicians are among the most respected professions on the planet, whose decision on the selected treatment may have a significant impact on someone’s life
7. Military Generals – Far beyond the field of war, military generals make far reaching decisions on a daily basis.
8. Bankers – They can be forced to issue a huge loan?

Landmark decisions in the future

Here are a few examples: Will our most important decision in the future be decided by well-informed individuals or a heavily biased A.I.?

1. Should cryptocurrencies replace national currencies?
2. Should we have a single world leader?
3. How should life and death decisions be made in the future?

Commandeered Systems

Every major system has the potential of being hijacked by an evil A.I. in the future. It can be achieved either through the tech itself, the people that control it or a combination of both, virtually all future systems will be vulnerable, such as:

1. Stock Exchanges
2. Power Plants
3. City Water Supply
4. Security Systems
5. Data Centres
6. Cloud Storage Systems
7. Airports
8. Prisons
9. Election Systems

Hijackable Equipment

As our equipment becomes more universally connected to the web, commandeered devices will become an ongoing concern. For example, the same drone that can deliver packages can also deliver bombs, poison, and spy on your kids.

1. Flying Drones
2. Driverless Cars
3. Airplanes
4. IoT Devices
5. Delivery Trucks
6. Stoplights
7. Smart Houses

Hard weaponized AI

South Korea currently maintains the border with its northern neighbour using Samsung-built robot sentries that can fire bullets, so it's safe to say autonomous weapons are already in use. It's easy to conceive future versions that could, say, use facial recognition software to hunt down targets and 3D-printing technology that would make arms stockpiling easy for any terrorist. Robotic soldiers would only aim at specific targets. They will be so small and cheap that even an average earner (say a potential terrorist) could buy it.

However, an individual robotic soldier would not be a threat to Humanity. What may create an existential risk is a potential arms race in autonomous weapons and Artificial Intelligence. Such a race would expose civilians to undue, potentially existential risk. If autonomous weapons are developed and deployed, they will eventually be in the air, space, sea, land, and cyber domains.

Paulo Santos writes in the Bulletin of the Atomic Scientists that such robot-soldiers will be taught to operate in teams, supported by a network of unmanned weapons systems. They will also patrol computer networks and potentially will

be everywhere. It is highly unlikely that only one country will pursue their development. As mentioned earlier in this section, Russia, USA, and South Korea have already displayed their capabilities. Thus, many states will conclude that they require development of ever-stronger artificial intelligence controlling various weapons with ever greater autonomy.

However, the main existential threat is this. **Autonomous systems with learning abilities could quickly get beyond their creators' control.** They would be a danger to anyone within their immediate reach. And autonomous weapons connected to each other via networks, or autonomous agents endowed with artificial intelligence and connected to the Internet, would not be confined to a single geographic territory or to states involved in armed conflict⁽⁵²⁾.

The unintended effects of creating and fielding autonomous systems might be extremely severe if they get under the control of malicious AI agents. In the worst-case scenario, nuclear war heads may be fired, almost definitely annihilating most life on earth, should all current nuclear arsenals be used.

Mitigating the risk of Weaponized AI or cyber soldiers

Thousands of signatories, including Stephen Hawking, Steve Wozniak, and Elon Musk, have lent their support to a letter recommending care and caution in the development of autonomous weapons. The letter, organized by the Future for Life Institute, was originally announced on July 28 2016. The Future for Life Institute is dedicated to mitigating the potential risks in world-changing technologies such as artificial intelligence, and has penned cautious recommendations about developing AI.

The letter argues against starting an AI arms race, preferring that AI be used in the military only to make battlefields safer for both military personnel and civilians: “Just as most chemists and biologists have no interest in building chemical or biological weapons, most AI researchers have no interest in building AI weapons — and do not want others to tarnish their field by doing so, potentially creating a major public backlash against AI that curtails its future societal benefits.”

The Institute recommends that the type of international agreements and treaties governing chemical or space-based weapons be also applied to AI, preventing use of autonomous weapons on the battlefield by any nation.

In the end, the most effective tool for fighting weaponized A.I. will be a more powerful benevolent A.I. This may create an arms race, as during the Cold War.

Nanotechnology

Nanotechnology would allow in principle to arrange atoms in any way we want. If this comes true, the world will forget what scarcity means because everything is made of atoms. This will move us into the future of abundance. As the philosopher Jason Silva puts it in “Future of Everything” “It essentially makes the physical world a programmable medium.”

One such scenario includes building the so-called nano-factory, a hypothetical device that could manufacture products with absolute atomic precision for a fraction of the cost of current manufacturing (53). “Atomic precision” here means that two objects produced by a nano-factory, for example two computers of the same design, would be identical with respect to not only their macroscopic properties, but also the precise placement of their constituent atoms. It remains unclear whether nano-factories are physically possible, but if they are, as theorists like Eric Drexler of the Future of Humanity Institute and Ralph Merkle of Singularity University claim, then the consequences for humanity would be profound.

There are only three resources required to operate a nano-factory: power, design instructions and a simple feedstock molecule such as acetone or acetylene. With these three conditions met, terrorist groups and lone wolves of the future could potentially manufacture huge arsenals of conventional and novel weaponry, perhaps eluding detection by law enforcement or international regulatory bodies. Nano-factories might even be capable of making nuclear weapons, although at present this possibility is uncertain.⁽⁵³⁾

Molecular manufacturing requires significant advances in nanotechnology, but once achieved it will be possible to produce highly advanced products at low costs and in large quantities in nano-factories of desktop proportions. When nano-factories gain the ability to produce other nano-factories, production may only be limited by relatively abundant factors such as input materials, energy, and software. Being equipped with compact computers and motors these could be increasingly autonomous and have a large range of capabilities⁽⁵⁴⁾.

However, ultra-precise manufacturing on an atomic scale, which could create materials with wonderful new properties, could also be used in frightening new weapons. There is even a possibility to create self-replicating nano-machines taking over the planet. Phoenix and Treder classify catastrophic risks posed by nanotechnology into three categories:

- From augmenting the development of other technologies such as AI and biotechnology.

- By enabling mass-production of potentially dangerous products that cause risk dynamics (such as arms races) depending on how they are used.
- From uncontrolled self-perpetuating processes with destructive effects (54).

This last speculative threat involves the intentional design of autonomous “nanobots” that would convert all the matter in their vicinity into clones of themselves. The result would be a positive feedback effect that could destroy the entire biosphere in as little as 90 minutes, according to a 2006 calculation by Ray Kurzweil in his book “The Singularity Is Near: When Humans Transcend Biology” (10). You may disagree with this prediction, but many existential risks experts take it seriously.

If you cannot imagine self-replicating nanobots, the size ranging from less than a millimetre to a few millimetres then have a look at a snowflake which is replicated from water molecules each time the weather conditions are right. But perhaps even a better example is your DNA molecule. Human DNA consists of over 3bn base pairs contained in 23 chromosome pairs. The thickness of the pages on which I am typing this text is roughly equal to a human hair, i.e. about 100,000 nanometres. But a strand of human DNA is just 2.5 nanometres in diameter, about 40,000 times thinner. In most simple terms, DNA has a property to replicate itself, modify and repair errors but in the end, it works like an unusual computer code, which always makes a copy of itself. It uses the original like one part of a ‘zip’ to create the second part of the ‘zip’. If it is placed in an appropriate growing environment, it creates billions of proteins, then cells, and tissues until it ultimately creates a living organism, which can then again reproduce. This is biology at its best.

However, a similar process can be developed for a non-biological environment, where atoms are bound into molecules, say mainly based on silicon rather than carbon, forming non-biological building blocks, an equivalent of human cells, which have been originally programmed to become a very small micro motor, still retaining the capability embedded in that motor, of replicating itself and creating billions of such motors ad infinitum. Such nanobots would behave like cancer and would ‘eat’ anything they could in order to create more copies of themselves or built new objects based on the instructions either retained in their ‘DNA code’ or received via Internet cloud. They could of course be affected by the same principle of evolution as a biological DNA and after a few years, produce objects that were never intended. And that unpredictability is the biggest danger.

Some experts in nanotechnology risk field suggest that the existential risk from that area comes from reaching technological advantage in arms race through the

availability of nanotech weaponry, which may destabilize a relative current balance between major powers. They list several reasons:

- A large number of players may be tempted to enter the race since the threshold for doing so is low;
- The ability to make weapons with molecular manufacturing will be cheap and easy to hide;
- Therefore, lack of insight into the other parties' capabilities can tempt players to arm out of caution or to launch pre-emptive strikes,
- Molecular manufacturing may reduce dependency on international trade, a potential peace-promoting factor;
- Wars of aggression may pose a smaller economic threat to the aggressor since manufacturing is cheap and humans may not be needed on the battlefield.⁽⁵⁴⁾

In addition, advanced nanotechnologies could introduce new nanoparticles to the biosphere, some of which could prove extremely toxic.

Mitigating Nanotechnology and experimental technology accident

The good news is that this technology in its extreme form is unlikely to be available this century. However, some crude and potentially equally dangerous nanobots could be available within a few decades. One of the key measures counterbalancing nanotechnology risk could be self-regulation by all state and non-state agencies. But that would be very hard to achieve. Instead the creation of an international organization similar to the International Atomic Energy Agency IAEA has been proposed. The Center for Responsible Nanotechnology suggests some technical restrictions and improved transparency regarding technological capabilities to facilitate arms-control in this area⁽⁵⁵⁾.

However, a resolution to this problem can already be found in nature. Bacteria, which are biological nanobots, have their own anti nanobots – the antibodies produced by our immune system. Therefore, nanobots which would behave dangerously could be wiped out by counter nanobots, assuming the scale of such a nanobots' 'epidemics' would not be too extensive and there would be enough time to develop those counter nanobots.

Laboratory Engineered Pandemics and Synthetic Biology

Biotechnology can pose a global catastrophic risk in the form of natural pathogens or novel, engineered ones. Such a catastrophe may also be brought about by usage in warfare, terrorist attacks or by accident. Terrorist applications

of biotechnology have historically been infrequent. To what extent this is due to a lack of capabilities or motivation is not resolved. It is believed, that exponential growth has been observed in the biotechnology sector and some scientists (Noun and Chyba) predict that this will lead to major increases in biotechnological capabilities in the coming decades. They argue that risks from biological warfare and bioterrorism are distinct from nuclear and chemical threats because biological pathogens are easier to mass-produce and their production is hard to control (especially as the technological capabilities are becoming available even to individual users).

More risks stemming from novel, engineered pathogens can be expected in the future. Scientists suspect that there is an upper limit on the virulence (deadliness) of naturally occurring pathogens⁽⁵⁶⁾. But pathogens may be intentionally or unintentionally genetically modified to change virulence and other characteristics. One example of that is what happened to Australian researchers who unintentionally changed characteristics of the mouse pox virus while trying to develop a virus to sterilize rodents. The modified virus became highly lethal even in vaccinated and naturally resistant mice. The technological means to genetically modify viruses' characteristics are likely to become more widely available in the future, if not properly regulated ⁽⁵⁷⁾.

We should look at the danger of self-replicating synthetic, incurable viruses from a particular angle – the rogue researcher syndrome. One possibility is that a disgruntled individual might steal a virus and travel around the world releasing it. An important factor in the motives of such a person might be his religious or cult-like convictions that might, in his mind, justify the act (a mass murder, like ISIS, but on a global scale). This risk is even more significant if one considers that biotech laboratories usually have no provision for psychological profiling of their employees who could be either a lone disgruntled individual (i.e. perversely a member of security staff) or a laboratory researcher.

Genetic engineering of new super-organisms could be enormously beneficial for humanity. But it might go horribly wrong, with the emergence and release of an engineered pathogen, accidentally, or through an act of war, targeting humans, or a crucial part of the global ecosystem. The impact of such an attack could be even worse than any conceivable natural pandemic.

Mitigating engineered pandemics and synthetic biology

Previously mentioned authors Noun and Chyba⁽⁵⁷⁾ propose three categories of measures to reduce risks from biotechnology and natural pandemics:

1. Regulation or prevention of potentially dangerous research

2. Improved recognition of outbreaks
3. Developing facilities to mitigate disease outbreaks (e.g. better and/or more widely distributed vaccines)

I would also add psychological profiling and very thorough investigation of the background activities and lives of these scientists, horrible as it must seem to anyone. After all, that is what any state does to its spies, in the name of a 'greater goodness' of the society.

Chapter 4

How Can Risks Combine into Existential Threats

How global catastrophes can become existential risks?

In the previous chapter I have covered existential risks, where each of them individually could wipe out Humanity. But there are risks which are global catastrophes but not existential on their own (i.e. humans may survive such an event). However, because of their complexity and interconnection with other risks they may become existential through their **combinatorial** effects, i.e. if they are triggered off at the same time with other risks, about which we may not be even aware of today. Such risks are the effect of what I call Global Disorder.

Most of us, if we look carefully around, see plenty of symptoms of such a mega crisis on the horizon. In most general terms, this category covers mismanagement of global affairs so serious that it may become the primary cause of our civilization's collapse, especially when combined with other risks. This includes global migration on unprecedented scale caused for example by prolonged draught in Africa, ensuing famine, and civil wars. Europe has already experienced a very mild migration of that kind in 2015-2016.

But global socio-political disorder may actually be more acute in the northern hemisphere, in the more advanced part of our civilization, for different reasons than famine. The origins of the social unrest will be in the collapse of the basic structures that underpin the western civilization, such as democracy, capitalism, the concepts of freedom, equality and responsibility, the ultimate fall of religion and the associated values. This may lead to economic and societal collapse, involving civil unrest and a breakdown of law and order that might make the continuation of civilised life impossible anywhere on Earth.

The risks of Global Disorder are linked to three domains:

- **Social Disorder**, created mainly by the unresolved problem of social inequality and wealth distribution, intolerance, and the impact of accelerating change, for which societies are totally unprepared
- **Economic Disorder**, created by the economic instability and a disconnection between the real economy and wealth creation that has led to the crisis of capitalism

- **Political Disorder**, which is caused because of the crisis of democracy and the absence of a credible global organization that could act as a kind of the World Government, being an arbiter in political and military conflicts

Of these three disorders, the Political Disorder is probably the most crucial because if that could be resolved, the other two disorders would have also become far less risky. The starting point for mitigating that risk successfully would be the creation of a new global organization with extensive powers, as well as a new system of governance based on a renewed global social contract. In my view, only such an organization may have a chance to save Humanity. The rest of the book covers how that might be achieved.

Global Social Disorder

This book is mainly about the future of Humanity. It is, therefore, appropriate to look at ourselves how we behave as a group in communities and societies. When we look back at the last century and realize what we as humans did to other humans, the conclusions one must draw is that at a fundamental level not a lot has changed in comparison with the previous centuries. There is still a lot of beast in many humans. Why is that?

Perhaps the best shortcut answer to that question can be found in the Maslow hierarchy that I wrote about in chapter 1 when discussing our civilization. What that hierarchy shows are that our actions and decisions are first of all driven by two factors: physiological needs and safety. Now imagine an average person voting in the UK Brexit referendum and look at the arguments that I will present from that perspective. Politicians appealing to Remainers and Brexiteers used exactly the same arguments – talking about **economic needs** (the UK would save £10bn net), **safety** (increased level of sovereignty, and migration control, by controlling the borders). Very few politicians addressed higher needs in Maslow's hierarchy towards the top of the pyramid e.g. **belonging** (similar democracy, history, friendship) or **self-actualization** (fuller potential and creativity – e.g. Erasmus programme, CERN, European Space Agency, tourism, etc.).

The point I am making is that to survive we must eat and be safe and that has remained man's main preoccupation for millennia– an evolutionary trait, a Darwinian struggle for the survival of the fittest. The fact that our GDP per head is about 20 times higher in real terms than 200 years ago does not change anything. And then there is the need for safety. It is born purely out of fear and of course has an emotional underlying. Fears drive our **emotions** and mostly create instantaneous responses, like Europe-wide anti-emigration protests over

the last few years. Deep down, at a physiological level people have broadly the same needs and fears.

The consequence of that is that the scope for making **rational** decisions in politics **for vast majority of people is very narrow, practical, and must include achievable results over a very short period.** The last point is perhaps best illustrated by the recent calls by the Brexiteers in the UK ‘to get on with it’, meaning the Government should get out of negotiations right now. They just cannot wait any longer for the result of the negotiations, almost irrespective of the consequences.

The post-war liberalism, especially in Europe, was an antidote against most brutal oppressions during the Second World War. It was necessary for Humanity to regain its faith in its capability to renew itself and be motivated by our most inspiring inner qualities. From that perspective, the post-war period has been just amazing, and it still is today. However, there is also another perspective.

The focus on human rights alone has been tilted so much that it has led to irresponsible behaviour on a massive scale. The state has been perceived as ‘they’, who ‘have’ (whatever that ‘have’ means at any given time) and as the one whose main obligation is to fulfil the needs of those who ‘do not have’. Because those who ‘do not have’ feel it is their right ‘to have’, and please do not ask who will pay for it. We have completely forgotten about the reciprocity and responsibility for our own lives and the lives of those, who sometimes for no fault of their own, cannot cope with the adversities of life.

President John Kennedy saw that danger very clearly when he famously said in his presidential inaugural speech on 20 January 1961: “Ask not what your country can do for you, ask what you can do for your country”. To turn these words into action, he set up the Peace Corp, a predecessor of current hundreds of voluntary organizations. One may argue it was a riposte to the Soviet Union’s export of tens of thousands of experts and technicians to developing countries to spread communism under the cover of help. However, I would look at it from a more benevolent perspective – the focus was on young people to help anybody truly voluntarily anywhere in the world.

Today, most of young people concentrate almost exclusively on their own needs rather than thinking of what they could do for members of their communities. Many of them do not even participate in elections. Some of them rely almost exclusively on the state’s handouts and then blame the ‘elites’ for not providing enough. But it was not always like that. We have already forgotten that for at least 40 years of the post-war period in most democracies young people served for a minimum of one year in the military service. Here the benefit for a society was two-fold. Young people were taught that freedom did not come free and it

was their responsibility to defend their country in the hour of need. But the second long-term benefit was even more important in peace time. They had a period of maturing into people who considered responsibility in a more general sense as the other side of the equation, balancing their rights.

Since the early 80's that system seemed to have collapsed. The way young people behave today was at least partially due to abolishing the military service or the mandatory public service, as is still the case in Germany and now again in Sweden. But it was also due to utter failure of the education system, the increase of wealth, more free time, and the ability to become self-sufficient (thanks to the generosity of the state). When I write about 'young people' I do not mean just today's generation of teenagers but also two generations past. Some of those young people who were brought up according to that model of state generosity and minimal responsibility for societal affairs, are now well into their middle age. So, today the vast majority of people in western democracies are completely oblivious to a simple fact that their citizens' rights must be balanced with their responsibilities.

That has obvious consequences. The state, keen to fulfil our immediate wishes, also stopped being responsible. The lack of the state's responsibility for our long-term well-being can be seen in many western countries, such as the USA or the UK. It manifests itself into pretending that everything is generally OK and literally patching up any problems and hoping that they do not come out just before the next election. Such a literary example is filling the potholes in the road after the winter rather than laying a new surface, which would have been a more durable and in the long-term and more cost-effective solution. However, for a local council it has some sense, because it can say its budget is in perfect order, people pay smaller taxes, so please re-elect us.

It is this kind of irresponsibility of not telling the nation the truth because that is not what the electorate wants to hear. It is even difficult to blame governments for this – this is simply showing who we are as species (the Maslow's hierarchy again). We are motivated in most cases by our most evolutionary instincts – meeting immediate physiological and safety needs. As an example, take the percentage of personal savings. In the UK it is very low at about 5% in 2018 down from 8% in 2015 (in Germany it is up from 8% to 10% in 2018). The hope in the UK is that the rainy day never comes. The reason people do not save as much as they used to in rich countries is because they feel they can always count on the government. Perhaps it is the feature of humanity as a species that the better off you are the less prepared you are for the worst? The fact that we do not save and do not take precautions as we have done before is a more general symptom of shedding our responsibilities onto others. We have become an irresponsible species!

So, here is an example of how ill prepared we are as a society to withstand even the slightest departure from what is considered a ‘normal’ life. It happened in London in wintery March 2018. There was a 5 cm of snow fall and the minimum temperature over two days was -4C in the night. But the result was that most trains stopped running, most schools were closed, and tens of thousands of people were without water for a few days, because main pipes broke down. Snow and frost in London are rare and happens on average for about a week every few years. But transport and social life is completely disorganised. As far as I am aware at the same time in Europe, where the temperature fell to -20C with heavy snow falls, the repercussions were far less severe. Yes, there may be some societal differences how various countries prepare for a proverbial ‘rainy day’. However, irrespective of that such an incident illustrates much bigger problems with fighting risks of any kind anywhere, in particular in western societies because it is a question of certain attitude to how societies manage their life.

If we look at such events as above from a wider perspective, it is clear that politicians prefer not to tell the voters the whole truth about the state of their country’s affairs, because most people prefer not to hear it. So, politicians follow the voters’ instincts, or preferences, if you like, i.e. people are interested only in what they can get here and now. Therefore, any long-term projects like infrastructure investments are not going to win many votes. That’s why there is such an unbelievable underinvestment in the American infrastructure, about which even Donald Trump lamented. This is a common situation in most countries.

So, what kind of future will we have in the age of Superintelligence? Perhaps we can find it in the quotes of these two famous science fiction writers. The first one, already mentioned, is Isaac Asimov – an American science fiction writer who characterised society as follows: “The saddest aspect of society is that science gathers knowledge faster than society gathers wisdom”. Somewhat strikingly, Stanislaw Lem – the Polish philosopher and science fiction writer offers a similar assessment when he says: “The fact that competent experts must serve under politicians of mediocre intelligence and little foresight is a problem that we are stuck with.” That is one of the reasons why I would propose a much clearer delineation between the legislative and executive powers, leaving the art of governing a country to real experts. That may become much easier with the acceptance of the next generation of AI Assistants in the next few years.

One of the areas where societal disorder can increase beyond control is the media, including social media and the Internet. The best example for me is the interview of the former Prime Minister Tony Blair during the Leveson Inquiry in the UK Parliament on 28th May 2012. Here is an extract from that interview as reported by BBC:

“Former Prime Minister Tony Blair has defended his friendship with Rupert Murdoch, saying ... a close relationship was inevitable but also involved a ‘certain level of tension’. After all, in this day and age, it is ‘essential and crucial’ to have good relationships with the media. With any of these big media groups, if you fall out with them, then watch out, because it is relentless. You then are effectively blocked from getting across your message’. In his witness statement, Blair said that media owners use newspapers and other media **as instruments of political power.**”

And that was said by the politician who was aware that Mr Murdoch’s company News Corp controlled almost 40% of the UK media, including the newspaper “The Sun”. The situation in the USA and its impact on politics is similar, where Comcast controls almost 40% of all media, including more than half of the broadband market. In 1983, 90% of US media was controlled by 50 companies; as of 2011, 90% was controlled by just 6 companies.

Global economic disorder

What’s wrong with Capitalism?

The capitalist system has served us quite well, at least by creating an incredible civilisation and increasing GDP per head in Western Europe on average 20 times in the last 200 years. But over the last 40 years it has become largely dysfunctional mainly by creating an enormous difference in income between the ‘capitalist elite’ and the rest of us (more on that soon) and gradually destroying democracy. So, what I want to do is to list just some of the most cardinal errors in the capitalist system that drive the world economy astray, and then select one or two areas that are most closely related to the subject of the book.

1. **Concentration of economic and political power** on a global scale by a tiny proportion of the world’s population. Corporate capitalism has been criticized for the amount of power and influence corporations and large business interest groups have over government policy, including the policies of regulatory agencies and influencing political campaigns. Many social scientists have criticized corporations for failing to act in the interests of the people; they claim the existence of large corporations seems to circumvent the principles of democracy, which assumes equal power relations between all individuals in a society⁽⁵⁸⁾.
2. **Corruption of democracy.** This is perhaps the most serious charge. Watching how the Donald Trump campaign was run, one might broadly agree as this is not only the symptom of failure of capitalism in the USA, but in most western democracies, e.g. Japan, Israel, Italy (Berlusconi) to name just a few. It’s money that controls politics and not the other way around. In

- the USA it is more apparent than perhaps anywhere else because it is so public, like the 'deal' that candidate Trump made with the National Rifle Association for their monetary support. The astronomical amount of money spent on presidential campaign in the USA, now reaching \$1bn is nothing else but corruption in white gloves, since that money under the disguise of lobbying is then used to pay for favourable legislation e.g. no gun control.
3. **Extreme concentration of wealth.** One area where change is happening at nearly exponentially is extreme wealth concentration. According to Forbes, in 2017 the number of billionaires jumped 13% to 2,043 from 1,810 the previous year, the first time ever that Forbes has pinned down more than 2,000 ten-figure-fortunes. Their total net worth rose by 18% to \$7.67 trillion, also a record. The change in the number of billionaires - up 233 since the 2016 list - was the biggest in the 31 years that Forbes has been tracking billionaires globally. Gainers since last year's list outnumbered losers by more than three to one. ⁽⁵⁹⁾.
 4. **Capitalism's inefficiency.** Planned obsolescence has been criticized as a wasteful practice under capitalism. By designing products to wear out faster than need be, new consumption is generated. This would benefit corporations by increasing sales, while at the same time generating excessive waste. A well-known example is the charge that Apple designed its iPod to underperform after 18 months. Critics view planned obsolescence as wasteful and an inefficient use of resources ⁽⁶⁰⁾.
 5. **Lack of proper accountability.** Corporate managers are the de facto 'owners' of large corporations, since institutional shareholders, who represent real owners e.g. millions of pensioners, are simply individuals who are either former board members of similar companies or 'professionals' quite often going hand in hand with the board. That's why Annual General Meetings are so pointless because the board only very rarely is overruled by the shareholders, to the extent that there is a new trend called shareholders' activism attempting to give shareholders the real power they should have.
 6. **Instability and debt accumulation problem.** The main source for growth capital is not organic money, i.e. from the company's profits, or equity, i.e. selling part of company's shares, but bank debt. As long as a company can repay the interest, it is deemed to be solvent. Therefore, if it wants to take over another company, it increases its debt level, the only real guarantee being the ability to repay the interest. No wonder then that any significant fluctuation in money supply and interest rate can lead to severe turbulences in the markets, like during the 2008 largest financial crisis in history, directly caused by the collapse of the sub-prime mortgage market.
 7. **Subsidized industries.** Some examples are the airline industry (fuel), steel industry (electricity), oil and coal industry, farming, and agriculture in general.
 8. **The doctrine of maximizing shareholder value.** This is based on Milton Friedman's assertion that the so-called agency-based model puts the

obligation on the company manager (the agent) to conduct the business in accordance with shareholders' wishes, in precedence to other stakeholders. This is of course a shortcut explanation, as the matter is quite complex, but the main point is that such a model of managing companies has led to a significant abnegation of the traditional role of a company's owner, who had to take risks and responsibilities that shareholders do not do. The result of that has been that shareholders' rights, as owners of capital, have been far superior to other stakeholders, such as employees, who deliver intellectual and manual 'capital'. But most importantly, globally it has led to trading in companies on the stock exchanges like in commodities, leading to big economic and financial crises, like in 2008.

9. **The failing financial systems.** Banks and the financial sector in general have become widely criticized and blamed for how they behaved before 2008 financial crisis. 10 years have passed and barely anything substantial has changed. Banks don't do what they are supposed to do i.e. the economy's payment mechanism and an intermediary between savers and investors, providing capital to new and growing businesses. Instead, more often than not, some banks are gradually turning into casinos, investing not for the long-term in real economy but gambling on micro-second transactions and creating derivatives that represent the original value several times inflated. And then one day, we all pay for it, and the carousel spinning the money, goes on.
10. **The crisis of tax systems and the boom of tax heavens.** Some multinational corporations are well positioned to take advantage of tax havens, but smaller businesses, are unable to do the same. That already creates an uneven playing field and distorts competition. Tax revenue is essential to fund vital public services such as education, health, and infrastructure, etc. By hiding income and assets, tax havens allow such companies and unscrupulous individuals to evade taxation, thereby allowing them to amass even more wealth and making inequality worse. Perhaps more importantly, the secrecy promoted by tax havens makes corruption pay, and with impunity. A conservative estimate of assets hidden in 'tax haven' jurisdictions is \$7.6 trillion in 2013. New estimates of revenue losses made in 2015 by International Monetary Fund estimate total tax losses at approximately \$600 billion globally ⁽⁶¹⁾.

This is just a small selection of the charges laid at the door of several hundred years old capitalism. I am sure one could add quite a few more, but that is not the point. It is sufficient to see that capitalism is in real crisis, which can endanger not only the economy but democracy and our freedoms. It has now reached its "sell by" date, and its basic principles need to be redefined.

Mitigating Economic Disorder Risks

To counterbalance the danger of the current form of seriously distorted capitalism, which may lead to an existential risk by combinatorial effects with other man-made risks, such as climate change, large scale migration, and world-wide conventional war, we need some fundamental changes in this system. We should go back to the drawing board and question some fundamental untouchable rights that underpin capitalism but also democracy. The revision of those rights needs to be supported by a new legal system, which initially will be impossible to implement world-wide but gradually it has to be rolled out globally. If this is started with a large economic area such as the EU, the pressure on other countries to follow the new laws set up there, could be exerted by severely limiting the trade with non-complying countries. The arguments that a single large country, such as the USA, cannot implement far reaching changes in its economic and trading system result that the worst aspects of globalisation are still with us. It can be done, but it needs a stiff resolve to withstand the counter pressures exerted over a long time. Here, are some of the proposals that I would think are realistic and could be completed within 10-15 years:

1. **Re-define what constitutes private ownership.** Shouldn't, for example, the right to land ownership be replaced with the right to rent land, since the land as other resources, should only belong to a country or a community?
2. **Put a cap on the maximum value of assets that can be held by an individual.** Shouldn't there be a cap for a maximum level of assets that an individual person can control? If this is not challenged, then very soon some individuals would control more wealth than some small countries. How far can we let it go? Couldn't, for example, limiting personal wealth, including inheritance, to \$100M be a good start? I realize how earth shattering the introduction of such law would be. Where is the sacrosanct right to private property? Well, in my view, this is the price we may need to pay for a more economically and socially equal society, as well as to regain the control of our democratic institutions.
3. **Create a new set of laws that would change the relationship between the shareholders and other stakeholders** within limited and joint stock companies as well as the rights and obligations of all the stakeholders to each other and to the state. That would be very difficult indeed in a global economy that we have now but is a burning issue that will not go away. One of the implementation routes might be via G7 – the member countries control more than half of the world's economy. Some progress has been recently done in taxation, so this is good evidence that it is possible. Shouldn't all those that contribute to profit, have a share in it but in proportion to the wealth they have created? That would need a redefinition of the shareholders rights vs. the rights of all company's stakeholders to the accumulated wealth.

4. **Curtail global dominance of super-large companies**, by perhaps limiting their assets to a percentage of national annual GDP. Companies global expansion till 1970s' was curtailed mainly by the control of the amount of finance flow allowed and the way it could happen (in the legal sense rather than technological). Yes, I am aware of the negative impact of the world's GDP growth but the damage that uncontrolled global expansion creates and how it is then used to control politics, especially in developing countries, leaves no option but to control the way companies operate.
5. **Reform the law on the operation of stock exchanges**. One of the more drastic measures to be considered might be a deep reform of the stock exchanges that would specifically forbid takeovers and mergers, i.e. trading in companies rather in partial stock.
6. **Subsidize the renewable energy**. Governments should subsidize renewable energy until it starts competing with non-renewable energy sources without a subsidy. Renewable energy will become a major disrupting technology very soon. Most experts believe that in 2050 the prevalent type of energy might be the one that is not yet known.
7. **Regulate the Internet-based business**, while leaving its freedom to operate in non-business areas largely unchanged. The Internet business is the most dangerous form of business because of its ability to control our lives everywhere. The ownership and control of the Internet companies must be dispersed as must be their size. Countries, where such companies have their head office would have to be the executors of such a policy. Any internet business that is outside such a regulatory compliance regulation would be illegal world-wide.
8. **Strongly encourage the growth of so called 'Conscious capitalism' and 'Benefits companies'**. This is an entirely new way of how business can be created and operated. In 2015 there were 28 such large companies world-wide (62).
9. **Global financial transaction tax**. Implement and gradually expand a financial transaction tax of 0.1% on the exchange of shares and bonds and 0.01% on derivative contracts. This is the initiative originally proposed by the EU in 2013. The tax could raise 57 billion Euros per year if implemented across the entire EU.

Global Political Disorder

Like before the two great wars, the world again faces a series of most profound crises but this time they are categorized as existential risks, i.e. such risks "where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential" (63). Taking a big picture view, the chance of one of these risks materializing by the end of this century is between 25%-50%. We already had one such a "near miss" that could have

annihilated the entire civilization. That was the Cuban crisis in October 1962, which nearly started the first, and quite probably the last, nuclear war. That was an example of a political disorder risk, in this case the effect of the Cold War.

Before I discuss measures, which might make Global Political Disorder less severe, let me give you a very brief account on the results of the failure of diplomacy in the post-war period. Global Political Disorder has been with us for most of human history. WWII was the deadliest military conflict in human history, when about 60m people perished (about 3% of the global population at that time). But after the WWII we had at least 7 major wars (no. of victims in brackets): Korean War (1952-1953 – 1.2m), Vietnam War (1965-1973 - 3.8m), Biafra war (Nigerian civil war in 1960' – 3.0m), Iran-Iraq war (1980-1988 – 1m), Second Congo war (1998—2003 – 5.4m), Afghanistan war (2001- 2014 – 2m victims). There were also over a hundred military conflicts such as in Syria right now, with over 400,000 people killed, and civil wars like in Cambodia (1.7m victims) and Rwanda (1m victims), so that together the number of people killed in military conflicts after the Second World War is about 30m. Europe seems to have learnt a bloody lesson after the WWII and was an oasis of peace for over 70 years, apart from the conflict in Yugoslavia and more recently in the Ukraine.

There are quite a few potentially very serious military conflicts that are brewing up right now, and these are the essence of the current Global Political Disorder, such as Russia-NATO, China-USA, North Korea-South Korea/USA, India-Pakistan Israel-the Arab world etc. But I hope, and that is only a hope, however irrational it may be, that these potential super mega conflicts will somehow be frozen for a generation. The most optimistic expectation now is that we will be living in a period of the Second Cold War, this time including several nuclear powers.

Since I have already spent some time discussing nuclear war and there is much more on mitigating Global Social Disorder further on in the book, I will make this section short and focus on the long-term threats that not only sustain the Global Political Disorder but make it significantly worse. In my view, there are several fundamental reasons why we are having a Global Political Disorder, which may get much worse quite soon, if nothing of substance is done. These are:

1. **A crisis of Democracy.** I dedicate the entire Part 2 to this subject, so only a few comments here. That crisis regards the practice of democracy in advanced Western liberal countries and the very slow spread of democracy to countries that have never had that political system. The perceived failure of democracy in recent years, especially in the last 20 years or so, has been no encouragement for other countries to adapt this system. This crucially

- relates to China, which seems to be even more confident that its own autocratic system is superior to a western liberal democracy. That of course is a sheer propaganda, confirmed in February 2018, by removing the limit of the number of terms that the Chinese President can serve (used to be a maximum of 2 terms). For similar reasons Russia, which really only had a very short glimpse of something resembling the Western democratic system under Yeltsin, can point to failures of Democracy in the EU and in the USA.
2. **A total collapse of the United Nations**, especially its Security Council, which can only very rarely make binding resolutions because of the meteoric rise of China as a global power and dangerously overconfident Russia under Putin.
 3. **Terrorism**. This is a relatively new factor on the stage of global politics because very tiny groups, mainly Islamists, can hijack the global politics and enforce allocation of significant resources. Secondly, they are being used as valuable pawns in stirring up Global Political Disorder, e.g. in Syria, which has become a testing ground by major powers such as Russia, USA/NATO and other larger countries, such as Iran, Iraq, Saudi Arabia and Turkey.
 4. **The Twitter World**. This is a very new phenomenon, so skilfully used by Donald Trump. Its contribution to Global Political Disorder is that when it is used by a top world politician it can stir up political and social unrest on a global scale in a matter of hours. In most cases it can be ignored but if it is a prelude to some more sinister moves by a global power to distract or provoke the opponent, it may have serious consequences.
 5. **Cyber war**. This is how quickly science fiction becomes reality. That can be really dangerous either directly or indirectly. Directly, information gained can be used for unlocking nuclear arsenals. Indirectly, by getting access to most secret information, which can then be used for getting a political or military advantage and create a temptation for starting a war, having an initial advantage.
 6. **Transnationalism**. This started as "a new way of thinking about relationships between cultures", described by Randolph Bourne in the early 20th century. It is a social and economic phenomenon but could have severe political implications in the world of global mobile phone interconnected communities (64). The Economic transnationalism is commonly known as globalization. In this context, multinational corporations looking for minimizing costs, cross political boundaries and become cultural, economic and quite often political disruptors, with unforeseen consequences. Some of them can be positive but most are negative, especially if they stir up tendencies for unification in bordering regions of two different states. The best example of socio-political transnationalism is the current war in the Middle East and especially the objective of the Kurds to create their own state out of regions in four neighbouring countries: Turkey, Syria, Iraq, and Iran.

Chapter 5

Making the World a Less Dangerous Place

The Final decade?

This is probably the worst time for any futurist to make predictions. The main reason why it is so difficult to predict even the near future is that the scope, multi-dimensionality, and foremost, the pace of change is becoming almost seamlessly exponential rather than linear, which for a human mind is an unknown and a deeply disturbing experience. Perhaps one of the best examples of completely wrong scenarios for the future is the Francis Fukuyama's prediction of 'The End of History' in the early 90' – a super optimistic view of the new set of benevolent relationships that were to evolve between the superpowers, in the post-cold war period. How wrong he was! Therefore, 'The final decade' may not be that final after all. And yet, forewarned is forearmed and perhaps presenting the near future in a somewhat uncomfortable way may contribute to selecting most pragmatic solutions to carry us through this turbulent decade.

We can look into the future from two perspectives. The first one, most common, is to project our current experiences and the state we are in, taking the average pace of change in the preceding decades and centuries as a measuring stick, and assess in that way the probability of the future situation materializing. The other perspective is to visualise what the future may be in a decade or two, and then return to where we are now, assessing the probabilities of various scenarios. This is how I prefer to view the near future, 'the final decade', ending in 2030.

Here, you may rightly ask, what does the word 'final' really mean and why it is this decade that is to be final. In my view, this may be the final decade for humans when we can still control the way we want to live our lives, and ultimately how we want to evolve as a species. You can search renowned sources on the Internet to make your own judgement if the word 'final' is an exaggeration. But I will only use two key points, which in my view justify this assessment.

The first point is Climate change, when many, if not most, of the scientists agree that the tipping point of the global temperature increase will be reached by 2030, if sufficient measures have not been implemented by then. Failing to deal with

climate change properly will determine to a large degree the quality of our lives. Directly we may initially only feel discomfort, but gradually the climate change will become life threatening. Indirectly, and much sooner, it may turn into a powerful trigger for other existential risks, such as massive migration, wars, or pandemic, all leading to a global destabilization of political, military, economic and social balance. It is already too late for the current actions, including an imaginative EU's climate change budget, to halt the temperature increase by 2030. The only feasible way might be to start an urgent geo-engineering reset of the planet's climate, using temporarily the least environmentally damaging measures (and there are over 100 of them). So, we still have some control over climate change till the end of this decade. If we fail, the planet may become uninhabitable for humans by the end of this century.

The second one, much more difficult, is the threat of the AI development, which is also likely to be out of our control by 2030. To avoid any confusion, I am not talking about millions of robots taking over but rather a subtle, web-like self-learning, single entity, grown out of the current multitude of AI projects. If such an entity becomes malicious, then we may only have a few decades left before humans become extinct or enslaved. We can no longer stop that process, like we cannot uninvent a nuclear bomb. More importantly, whatever type of AI we will create, even if it becomes our friendly partner, we will be gradually handing over more and more control to it. Climate change, at least initially, will impact the quality of our lives. AI has been already impacting the decisions we make. Within this decade, we will be rubber-stamping at a personal and governmental level most of the decisions taken by an AI agent on the basis of a previously given consent. Today, we think we understand the consequences of these decisions, for which we have given a prior consent, and more importantly, we trust, that our smart phone that already embeds a lot of AI technology, is not outsmarting us for its own, potentially devious, hidden aims. But for how long will we be able to trust these agents, which already have some emotional features built in?

I realize it may be hard to accept such a dystopian view at this point because these scenarios are not on the front covers of our daily papers, at least not yet. However, let us assume that this is indeed the last decade when we can control our future. How do we see the current developments on the global political stage, the pace and breadth of actions taken, and the significance of the proposed solutions matching the above challenges? Wouldn't you agree that if this is our

‘final hour’, and if so, the actions of politicians of large and small countries, including the UN, are just incredibly insufficient? What actions then may be taken, or rather should be taken. I will start with the most desirable solutions from a global perspective before I come to the UK’s courtyard.

We cannot solve global problems, which can now be rehashed as Humanity’s problems, locally. They can only be solved globally. In order to do that, we need a global organisation with the powers of a federation, on par with the powers of the US government. We face two problems here. First of all, there is no way that such an organization can be created from scratch and achieve the desired effects in such a short time (e.g. reduction in a global temperature or having a global control over AI development). Therefore, it has to be an existing organisation, which would very quickly embrace not just its current members but also the majority of the countries on the planet. Realistically, there are only two organisations that might fulfil this role – NATO or the European Union. For various reasons, mainly the lack of diverse experience, NATO seems to be a less suitable organisation. So, the only remaining candidate is the European Union. However, there may also be a kind of the last resort candidate – China, with all the unwanted consequences.

If we assume, that the EU would take this role, it will of course have to become a federation and very quickly expand by including more and more countries (I leave aside a big potential problem that the EU may not want to do it). To play the role of the ‘saviour of Humanity’, the EU will not have to include all countries by 2030. It will be sufficient if it creates a critical mass, where its decisions will by default have to be implemented by most countries (the only current threat is banning the access to the EU market). That is what has happened with GDPR, which formally applies only to the EU, but actually is adhered to by the vast majority of the countries. There are also some very positive developments in the control of AI, with some EU’s legislation to be ratified by the end of this year. However, such a legislation requires literary a global control of AI and not just within the EU. In any case, the EU would have to become a de facto World Government by the end of this decade.

Now let us look what is actually happening in the EU and what are the latest initiatives in the area of EU integration (a codename for a federation). Many of the top EU politicians, like the German Chancellor Angela Merkel, heading the EU’s Presidency right now, or the former President of the EU Council, Donald

Tusk, have an idealistic objective of achieving ‘an ever-closer union’, i.e. federalization of the entire EU in one stage, even if it happens at a slow speed. They point out on the difference in the GDP level per capita, which for the top countries, such as Austria, Belgium or Germany is about 5 to 6 times higher than for the bottom ones, such as Bulgaria and Romania. They also say that waiting for another decade to start the preparation for the Federation, is a safer option because the currently poorer countries will make faster incremental improvements in their GDP and in the social area than the richer countries, making the differences between members states smaller, and the transition easier.

In my view, this should not be a preferred option and the whole EU should not be converted into a Federation at the same time. The main reason is that there will be new countries joining the EU, which will be less ready than the ‘old’ member states. Therefore, there may never be the right time for all the EU countries to federate together. Waiting a decade or more will be very risky indeed, see above. But on top of that, we now have a Covid19 pandemic, the effects of which may last throughout most of this decade. There are also, of course, all the ‘classical’ risks stemming from the increasing level of the world’s instability, e.g. the continuous meddling of Russia into the EU affairs and the volatility of the financial markets. The further the moment of the federation is pushed back, the higher the risk.

The EU Commission has an unapproved scenario 6 (courtesy of Mr Juncker), which envisages that EU should transform itself into the European Federation by 2025. Not so long ago it looked very unrealistic. But after Covid-19 pandemics, it will be a necessity. The post-pandemic economic and social crisis has already engulfed the EU, as it has the whole world. The ensuing chaos will be multi-dimensional including high permanent unemployment, caused by an earlier arrival of Technological Unemployment, significant drop in the GDP and social unrest. But there will be other areas that may simultaneously aggravate the super crisis, such as a pension crisis, or the rise of interest rates. Finally, there could also be other momentous events linked again to a sudden migration e.g. from the Ukraine, if it falls apart due to another coup d’état. This might lead to sealing off the internal EU border, the collapse of the Schengen zone and prompting some countries to exit the EU in a completely chaotic way. That could spark off the break-up of the EU into several groups such as the original six

founding members, the Visegrad Group, or the Scandinavian countries, which might re-join EFTA, where they were before.

To avoid that, the ‘core’ EU countries may be forced to start a quick fix, untidy, untested variant of the EU federalization, as a risk mitigation strategy. Germany might then see the creation of a Federation as a safer option both for Germany and Europe. Through the European Federation Government, it could take full control of the budget and thus reduce its potential losses. Therefore, crises like these may paradoxically be a trigger for a much faster federalization of the EU. This envisages a transition to a rudimentary federation, when only the most necessary functions would be federated, such as defence, security, foreign affairs, and the common currency. This is now, in my view, the most likely outcome.

Should such events happen in the next 2 years, then it could be president Macron, rather than Mrs Merkel, who might be the standard bearer for the EU federalization, almost exactly 200 years after Napoleon tried to do just that. What a perspective! In such a scenario, Italy might remain on the side-lines and play a less prominent role in the future federated EU, replaced by smaller but economically stronger countries such as Benelux, Ireland, and Sweden.

Whatever triggers the process of a Fast Track Federalization, the initiators may invoke Article 20 of the Lisbon Treaty. Like the article 50, which facilitated Brexit, it is ‘the article of the last resort’ or a kind of a ‘nuclear option’. That is why it is perhaps least known, because it has never been used. Additionally, it is one of the most complex and convoluted articles in the Lisbon Treaty, probably done in this way to hide the real intention of the proposers, since this is the gateway to a federated Europe. Therefore, I spare you the full wording of that article. In plain language it means that if at least 9 members of the EU want to federate, they can practically do it at any time. The post Covid-19 economic, social, and political upheavals may create unprecedented tensions within the EU. Therefore, if we include personal ambitions of some politicians, and deep structural budgetary problems of the Mediterranean countries (Greece, Italy, Spain, and even France), it may happen indeed within a few years.

There are of course other ways of achieving the Fast Track EU Federalization, which may even include countries currently outside the EU, namely the EFTA members – Norway, Switzerland, Iceland, and Liechtenstein. Paradoxically, if

the level of federalization is very shallow (which I think it should be), it might also become, at some stage, a face-saving solution for the UK. The main problem here would be the need for the UK to accept the European Court of Justice, and later on, the European Constitutional Court, as the supreme legislator. However, we live in such abnormal times that anything is possible. In this context, even the most advanced version of the proposed Conference on the Future of Europe, put forward by the EU Parliament, looks very uninspiring indeed.



The Conference is to discuss six subject areas (called Themes): Institutions, Economy, Security and Defence, Climate Change, Digital World and Social Issues. Each of these themes is to be deliberated in one of the six chosen EU countries at a series of debates at Citizens Assemblies over a 2-year period. The decision-making body is to be the Conference Plenary, but the European Council does not want these decisions to be binding and has not agreed to allow any members of the Citizens' Assemblies to take part and have a vote in the decisions made by the Plenary.

It is not yet known either, if this fresh attempt will fundamentally change how the EU works. Before the Covid-19 pandemic it was clear that a strong opposition would come from the usual suspects. These are the EU leaders – members of the European Council because they would lose most, should the

Conference end with a success, like the creation of the European Federation. However, what is encouraging, is a strong support coming from the members of the European Parliament.

Since the situation after the end of the current pandemic will be entirely different than just a few months ago, people supporting the European federalization, such as nearly a hundred federalist movements, should press for abandoning the current minimalistic goals. Instead the Conference agenda should be bolder and include the debates on the shape of the future European Federation, calling for a Constitutional Convention and setting a concrete date for the transfer of those EU members that want it, to the European Federation. They should thus use the momentum and the infrastructure of the Conference (with some modifications) but reject the proposed objectives as far inadequate from what is required right now. Instead, they should take Presidents Macron's call for giving more power to the EU Citizens literally. After all, if a representational democracy has failed so badly to meet the expectations of the electorate, the electorate itself must insist on formulating their demands in a direct form, e.g. through a network of Citizens' Assemblies.

The Covid-19 pandemic may have created another paradox. This time it relates to the United Kingdom. The changes on the British political stage have accelerated rapidly, although not necessarily in the right direction. What can be clearly seen from nearly a year of Boris Johnson's government, is an unrelentless push for the centralisation of power. People hoped that dramatic scenes in the British Parliament during the process of Brexit ratification, including the legally questionable proroguing of the Parliament, would end. Just to the contrary, as the latest attempt to nominate Chris Grayling as the head of the Intelligence and Security Committee by the Government, rather than by the Committee itself, proves. More importantly almost a provocative continuous disregard for the Scottish Government to be consulted on post-Brexit options, may only widen the current cracks in the system of the United Kingdom governance opening a real prospect of the end of the Union.

So, what is that paradox. In this short scenario there are many variants and 'ifs'. The first one is that the conservatives will have enough of Dominic Cummings and Boris Johnson. We can see an anti-ERG Group emerging in the Parliament, fuelled by Theresa May's bitter fight with Boris Johnson and also some dislike of his centrist policies by a growing number of Conservative MPs. That may

lead to an election of a new leader, such as Rishi Sunak, even before the post-Brexit deal is ratified. If it happens next year, which is more likely, Britain, Europe and the world will be in the midst of a real mega crisis, which when combined with the post-Brexit chaos may lead to a rapid resetting of the British political stage. This may include such momentous events such as: forming a government of national unity, avoiding the Scottish departure from the UK by having a Constitutional Convention, which will create a new British Federation.

Finally, this turbulent period may open a sudden opportunity for a quick re-entry to the EU, perhaps by participating in the Conference on the Future of Europe. This might allow Britain (again paradoxically) to have a significant say on the shape of the future European Federation, reflecting more the British view and indirectly justify Brexit a posteriori. This would mean creating a Federation, as a Minimal State, a miniFed, with only the most important areas being federated, such as defence, foreign affairs, the budget, and the common currency but built on the shared set of values.

Whatever happens, I believe this may be the final decade when humans are still in control of their future. But that does not mean that the future must be dystopian. Just to the contrary, we are only perhaps less than two decades away from the world of abundance. The problem is that we must cross a weak bridge over a bulging river that separates us from the opposite shore. There is a price tag attached to enable us a safe passage, which requires the resetting of our values and the relationship between the nations, so that humans behave like bees in a hive.

What are key strategies for mitigating existential risks?

If our civilization is to survive, we need to apply some powerful risk mitigation strategies. The 10 top existential risks listed earlier are total risks, i.e. they multiply probability by the risk impact. We have hardly any control over the non-anthropogenic risks. But we do have control over political, social, economic, and technological risks. That is why for our further consideration we will only evaluate mitigation strategies for anthropogenic risks.

All anthropogenic risks, apart from Superintelligence and climate change, are of a lottery type. They may or may not occur, e.g. global pandemic or a nuclear war. However, Superintelligence and Climate Change are the two types of risks, which certainly materialize, if we do nothing to mitigate them. In the case of

climate change, there is almost a certainty that if an average global temperature rises by more than 6C then the runaway scenario would be triggered, that may lead to total annihilation of human species. When that may happen? If we do nothing, such risk will most likely materialize in the next century.

The same is true about Superintelligence. If we do not prepare properly before the time when it arrives, then such risk will almost certainly occur in 20, 30 or latest 50 years. In any case, **it would arrive earlier than the risk related to climate change**. Therefore, it is the risk of Superintelligence is the most immediate and the biggest danger that Humanity faces. The remaining risks, such as pandemics or weaponized AI, nuclear wars, or nanotechnology, originate in technology related innovations or global social disorder but they are of a lottery type, so they may not happen.

There are not many realistic overall strategies proposing how Humanity could mitigate existential risks. Those that are available describe mainly policies which would need to be implemented if we want to combat ecological, social, and economic risks successfully. For example, World Economic Forum, in their otherwise excellent Global Risks Report, mentions mitigation of risks in just a few sentences over the entire 78-page report.

Similarly, Global Challenges Report 2016 that has one of the most advanced methods for identifying risk proposes the following 10 steps in mitigation of existential risks:

1. Global Challenges Leadership Network
2. Better quality risk assessment
3. Development of early warning systems
4. Encouraging visualisation of complex systems
5. Highlighting early movers
6. Including the whole probability distribution
7. Increasing the focus on the probability of extreme events
8. Encouraging appropriate language to describe extreme risks
9. Establishing a Global Risk Opportunity Indicator to guide governance
10. Explore the possibility of establishing Global Risk Organisation (22)

Establishing a Global Risk Organisation, which is the only concrete proposal in those 10 steps, which is a further evidence that risks reduction initiatives are partial and isolated, although in theory, if such an organisation had supranational powers then it would be a step in the right direction.

So, at the moment, any mitigation strategies are just simply a plaster to get the world through to the next year, as has been the case in politics of most nations, including western democracies. Short-termism is an outright winner. My only

hope is that today's media such as the Internet, may make more people aware of how serious the threat of the annihilation of all humans is. This could prepare millions of people for tough actions and sacrifices that may be needed, by understanding the severity of the problems better and in more realistic terms than just by reading the tabloids headlines. And that is also my intention – to make people more aware of what existential risks we face and what we can potentially do to avoid the ultimate disaster.

What is likely to happen if we just keep going on?

The biggest risk that our civilization faces is that nothing substantial will be done to combat existential risks. We do not have 100 years to prepare the mitigation strategies and allocation of the required resources. We may have just 10-20 years if things go really badly.

There is a point of view that civilizations resemble living things. They are born, have their immature youth, followed by a longer period of mature development until they finally reach the point of their collapse. There have been a number of significant near existential risks for our civilization. Think about the Thirty Years' War (1618-1648). That was fought only in Western and Central Europe when these nations went to war fighting for the right (in their mind) interpretation of some religious principles. It was, in relative terms, one of the longest and most destructive conflicts in human history where 8m people died. In Germany 1/3 of the population died and some towns were reduced to half of their original population.

It may be helpful to remind the warning given by Enrico Fermi, the Nobel Laureate in Physics on the consequences of not trying to mitigate existential risks. Fermi, apart from being one of the key people behind the construction of the first atomic bomb, also formulated the so-called Fermi Paradox. It is the apparent contradiction between the vast number of stars and galaxies in the universe and lack of evidence for intelligent life. Even if intelligent life occurs on only a minuscule percentage of planets around these stars, there might still be a great number of existing civilizations.

If the percentage were high enough, it would have produced a significant number of civilizations in our own Galaxy - the Milky Way. There are dozens of hypotheses explaining this paradox. But five of them directly relate to existential risks. These are:

- **It is the nature of intelligent life to destroy itself.** This is the argument put forward by Nick Bostrom in his book "Existential Risks - Analysing Human Extinction Scenarios and Related Hazards". It says that technological

civilizations may usually destroy themselves before or shortly after developing radio or spaceflight technology. This might have happen for a number of reasons such as catastrophic wars, environmental pollution, pandemic (if life is biological), or the artificial intelligence.

- **Self-annihilation.** This is an idea put forward by Stephen Hawking in "Life in the Universe". It says that life as such is an aspect of thermodynamics. As an ordered system, it can only sustain itself against the tendency to disorder by using energy. If that energy is insufficient, the system becomes unstable and self-destructs.
- **It is the nature of intelligent life to destroy others.** Another hypothesis, put forward by Steven Soter in "SETI and the Cosmic Quarantine Hypothesis" is that an intelligent species beyond a certain point of technological capability will destroy other intelligent species as they appear.
- **Periodic extinction by natural events.** This is an example of a natural risk. According to this hypotheses, new life might commonly die out due to runaway heating or cooling on their fledgling planets [75]. On Earth, there have been numerous major extinction events that destroyed the majority of complex species alive at the time; the extinction of dinosaurs is the best known example. It may be the case that such extinction events are common throughout the universe and periodically destroy intelligent life, or at least its civilizations, before the species is able to develop the technology to communicate with other species. (65)
- **Resource depletion and climate change.** This is an argument put forward by the astronomer Adam Frank who argues that industrial evolution on other planets may lead to a sustainability crisis, eventually leading to devastating climate change (66).



2

PART 2

SUPERINTELLIGENCE

Please note: You can find more detailed information on the subjects discussed in this Part, in Volume 3 of TRANSHUMANS – “Becoming a Butterfly”

Chapter 1

From Intelligence to Superintelligence

Developing Superintelligence

What is Superintelligence? Could it be our friend or become a foe, which in the worst-case scenario can annihilate Humanity? But before we answer these questions let us agree on what is intelligence? Unfortunately, this is not a term that is easy to define unambiguously. **Intelligence** can mean many things to different people. The scientific community has been debating this since at least the late 19th century.

Without going into a long discourse, intelligence is defined in a popular sense as a general mental ability to learn and apply knowledge in order to change the environment most effectively for the intelligent agent, in our case - humans. Recently, some scientists rejected the idea of a single intelligence and instead have suggested that intelligence is the result of several independent abilities, which when combined contribute to the total performance of an individual. That would include other “intelligences” such as:

- the ability to evaluate and judge
- the ability to reason and have abstract thoughts
- the ability to learn quickly as well as learn from experience
- the ability to comprehend complex ideas
- the capacity for original and productive thought

Robert Sternberg, a psychologist, proposes that there are three fundamental aspects of intelligence: analytical, practical, and creative. He believes that traditional intelligence tests only focus on one aspect – analytical – and do not address the necessary balance from the other two aspects. ⁽⁶⁷⁾

Let us now move to **Artificial Intelligence** (AI), which is a very new discipline. It has been applied in earnest for at least 30 years under various other names such as Expert Systems and later on Neural Networks. Its key features are super performance and imitation of human cognitive abilities like problem solving and learning or speech recognition. It can beat best human capabilities but usually in one discipline only. Therefore, it is termed as a “narrow AI”. In 2011, the IBM Watson computer system competed in Jeopardy game, against former winners Brad Rutter and Ken Jennings. Watson won the first place and the prize of \$1 million. Then in March 2016 Google’s AlphaGo computer using self-learning (machine learning) program beat the 18-time world champion Lee Sedol. The success of self-learning has sparked a real revolution in AI. Ray Kurzweil,

currently the Chief Technologist at Google and one of the most reliable futurists, (I will be quoting him a few more times), confirmed his earlier prediction that AI will match human intelligence (in narrow subjects) by 2029.

In the next 20-30 years, machine self-learning may deliver an intelligent agent, which will surpass any human being **in every skill** or task. When AI reaches such capability, it will become **Artificial General Intelligence** (AGI), or **Superintelligence**, the term used in this book after Nick Bostrom.

Very soon after Superintelligence has been embodied either in a robot or a computer network, it will be capable of redesigning itself. Imagine that such a smart machine will be capable of rapidly producing generations of progressively improved, powerful machines, creating intelligence far exceeding human intellectual capacity, until it reaches the so-called runaway effect ⁽⁶⁸⁾. Once Superintelligence has reached that point it may be impossible for a human to comprehend it and control it. It will thus quickly reach the point in time called **Technological Singularity or simply Singularity**. For the purpose of this book, **Singularity** means the point in time, when Superintelligence will through the process of self-learning become more knowledgeable and potent by re-inventing itself at an exponential pace.

Ray Kurzweil in his book “The Singularity is near” (10) predicts that Technological Singularity event will happen by 2045, while the SoftBank CEO Masayoshi Son, another authority on AI, forecasts it will happen by 2047. Ben Goertzel, one of the well-known AI researchers, who is chief scientist at the robotics company Hanson Robotics – the creator of Sophia robot, believes Superintelligence is possible well within Kurzweil’s timeframe. However, he also says that the Technological Singularity

To achieve Superintelligence, we need to make at least three major improvements. Tim Urban suggests how this could be done ⁽⁶⁹⁾:

- **Increase computer power.** This has been doubling every 18 months following the so-called Moore’s Law. Some people think the “Law” will stop working about 2030 because we will reach physical limits of continuing miniaturization of chips. On the other hand, some futurists, such as Ray Kurzweil, predict that by around 2025 intelligence packed into a \$1,000 computer, should reach the power of a human brain. But even that seems to be a fairly moderate prediction in view of exceptional acceleration in the development of a quantum computer (more on this later).
- **Emulate human brain using reverse engineering.** There are several ways to do it. One that Urban suggests involves a strategy called “whole brain emulation,” where the goal is to slice a real brain into thin layers, scan each

layer one by one, use software to assemble and accurately reconstruct a 3-D model, and then implement the model on a powerful computer. Recently we have been able to emulate a 1mm-long flatworm brain, which consists of just 302 total neurons. The human brain contains at least 100 billion neurons, each having on average about 10,000 connections. If that makes it seem like a hopeless project, remember the power of exponential progress.

- **Emulate human brain by copying the process mastered by evolution.** This could be done using a machine-oriented approach, not by mimicking biology exactly. To do that we would build a computer that would have two major skills: doing research on how it could improve itself and then coding changes into itself (that's exactly what evolution did to us). We would teach computers to be computer scientists, so they could bootstrap their own development. That would be their main objective, finding the most effective process to make themselves smarter.

But how could we measure intelligence? The current "gold standard" of measuring intelligence is **Intelligence quotient (IQ)**. When current IQ tests were developed, the median raw score of the people's sample was defined as IQ 100 and scores with standard deviation (SD) up or down were defined as 15 IQ points greater or less. By this definition, about two-thirds of the population scores are between IQ 85 and IQ 115. About 5 percent of the population scores above 125, and 5 percent below 75". (70)

Psychologist Howard Gardner's theory of multiple intelligences states that intelligence can be broken down into 8 distinct components: logical, spatial, linguistic, interpersonal, naturalist, kinaesthetic, musical, and intrapersonal. Thus, he believes that standard IQ tests and psychometric tests focus on certain components are not entirely reliable for assessing human intelligence. However, IQ tests are de facto the only standard that is widely used and despite its weaknesses it allows making some valuable insights into the intelligence and capabilities of millions of people. We may need to improve these tests but, in the meantime, they are already applied to intelligent assistants, such as Siri or Google's Personal Assistant.

On 3rd October 2017 a test was organized for several AI assistants by three Chinese researchers: Feng Liu, Yong Shi, and Ying Liu, primarily based on exams carried out during 2016. According to researchers, Google's AI Assistant rating of 47.3 is barely beneath a six-year-old human's IQ of 55.5. However, it was more than double that of Siri's IQ of 23.9. Siri is also behind Microsoft's Bing or Baidu, which have respective IQs of 31.98 and 32.92 respectively. All AI's IQs are considerably lower than a mean for 18-year-old's which is 97.

The researchers say, that: “The results so far indicate that the artificial-intelligence systems produced by Google, Baidu, and others have significantly improved over the past two years but still have certain gaps as compared with even a six-year-old child” (71). They grade AI’s intelligence into six levels based on the model that combines AI and human traits around four areas of data, together with “input, output, mastery, and creation”:

- First-grade system, which might exchange some information with people
- Second-grade system that can manage the interface to some objects such as TVs or washing machines, the so-called Internet-of-Things (IoT)
- The third grade includes computer systems and mobile phones, which are programmed and can be upgraded. That would include AlphaGo from Google’s DeepMind
- Fourth grade include Google Brain, Baidu Brain, and the EU’s RoboEarth robots, because they have the ability to communicate and be managed using cloud data
- Fifth-grade intelligence is at a human level
- Sixth-grade system will have the capability to “continuously innovate itself and create new knowledge, with I/O ability, knowledge mastery, and application ability that all approach infinite values as time goes on” (71).

The difference between the grades of AI seems to be quite significant. But once it gets to the sixth grade, AI will improve exponentially until Superintelligence becomes Technological Singularity.

How soon can it happen? There is no agreement on when Superintelligence may definitely arrive. The Future of Humanity Institute at Oxford University did a research in May 2017 asking 1,634 researchers who published papers at the 2015 NIPS and ICML conferences (the two leading machine learning conferences) and asked them to complete a survey on when AI will outperform humans in various areas. 352 researchers responded and their aggregate view is as follows (Asian respondents expected these dates to be much sooner than North Americans)⁽⁷²⁾:

- By 2024 - translating languages
- By 2026 - writing high school essays
- By 2027 – driving trucks
- By 2031 – working in retail
- By 2049 –writing a bestselling book
- By 2053 – working as a surgeon
- By 2062 – 50% chance of AI outperforming humans in all tasks

Perhaps the best-known predictions are those made by Ray Kurzweil, one of the prominent futurists, who has already proven over the previous 30 years that his predictions were largely correct. In an interview Kurzweil had with *Futurism* on 3rd October 2017, he confirmed that: “2029 is the consistent date I have predicted for when an AI will pass a valid Turing test and therefore achieve human level of intelligence. I have set the date 2045 for the ‘Singularity’ which is when we will multiply our effective intelligence a billion-fold by merging with the intelligence we have created.” For Ray Kurzweil, the process towards this singularity has already begun⁽⁷³⁾. If you think that Singularity cannot happen by 2045, then these two latest inventions from Google should make you think.

First, was the invention announced by Google at its I/O conference in May 2017, which gives quite a considerable boost to those impatiently awaiting the arrival of Superintelligence. It was its latest invention – AutoML (Auto Machine Learning). The Google team came up with a machine learning software that can create self-learning code. The system runs thousands of simulations to determine, which areas of the code can be improved. It then makes the changes and continues the process until its goal is reached. The result was that AutoML is **better** at coding machine-learning systems than the researchers who made it. In an image recognition task, it reached record high 82 percent accuracy. Even in some of the most complex AI tasks, its self-created code is superior to humans⁽⁷⁴⁾. Google’s AutoML could develop a superior image recognition system within a few weeks, something that would have taken months for most brilliant AI scientists. But in December 2017, just six months later, the Department of Energy’s Oak Ridge National Laboratory (ORNL), using the most powerful supercomputer in the US, have developed an AI system that can generate neural networks as good, if not better, than any developed by a human **in less than a day**.⁽⁷⁵⁾

The second invention, announced in the journal “*Nature*” on 19/10/2017, will probably be viewed in the future as a very significant milestone on the path to Superintelligence and Singularity. Barely a year after Google’s AlphaGo beat Lee Sedol, the Grandmaster in the Chinese game GoGo, which itself was considered a very important breakthrough in AI, a vastly superior AI agent called AlphaGoZero (AGZ) beat GoGo Masters by 100 games to 0. The original AlphaGo had the benefit of learning from thousands of previously played Go games against human players and against itself. AGZ, on the other hand, received no help from its human handlers, and had access to absolutely nothing aside from the rules of the game. The only input it had were the rules of the game. Using “reinforcement learning,” AGZ played against itself 4.9 million games, starting from a very basic level of play without any supervision or use of human data (AlphaGo, by comparison, had 30 million games). The self-learning capability allowed the system to improve and refine its digital brain, known as a neural network, as it continually learned from past experience. In effect, AGZ

was its own teacher. It took for AGZ just **4 hours** to self-learn chess to such a degree that it beat the world class champions. Additionally, this technique is so powerful because it is no longer constrained by the limits of human knowledge. Instead, it can learn from a “clean slate”.

The way in which AGZ teaches itself is so significant for AI because it shows that once AI gets full knowledge about the real-world problem to be solved then the power of reinforcement learning will deliver the result. Gradually such AI will become Artificial General Intelligence (AGI) i.e. Superintelligence, finding solutions and strategies that are beyond human capabilities. The most surprising in this story is how short the time was between the AlphaGo’s win and an absolute supremacy of AGZ over GoGo Masters. That’s what exponential progress is about.

Superintelligence may potentially become the most dangerous adversary of Humanity. We shall devote more space to that eventuality later on when we discuss Superintelligence as Humanity’s greatest existential risk. However, it can also become very benevolent and friendly agent that Humanity will badly need to fight other existential risks and help turn our civilisation on the path of immense prosperity. Let me then begin with describing that possibility first.

From Superintelligence and beyond

Most of you reading this book will live to see the advent of Superintelligence irrespective of it being conscious or not, provided that we do not make ourselves extinct before then. In that context, we need to get the sense of what is the very nature of Existence and how Superintelligence can evolve as a potentially distinct species far into the future, aging within the aging Universe. But for the purpose of extrapolating the evolution of Superintelligence, we need first to go back and look how we, humans, came into Existence from Nothingness.

It is my very personal, perhaps a bit unusual, view of Nothingness and Existence. I remember well when in 2005, after a lecture at the London’s Royal Society of Arts and Commerce (RSA), I spoke to the Nobel laureate in chemistry Harry Kroto. I asked him "How would you define 'Eternity'?" He was a bit surprised but then asked me if I had such a definition. I proposed this one: "Eternity is Nothingness on average", with which he did not disagree, since there are a number of theories that come to similar conclusions like the Dynamic Eternal Universe (76). I shall now use this as a starting point and list, in a simplistic way, the steps that led to the creation of humans.

There are a number of theories, which describe how the evolution of the universe has begun. The best known is the theory of the “Big Bang”. The model of our

Universe since the Big Bang is supported by many detailed, and in most cases, proven theories, such as quantum physics with its Standard Model, relativity, gravity, and the strings theory. There is still, however, no overall theory called the Theory of Everything that could reconcile gravity and quantum theory. Theories that try to explain what existed **before** Big Bang, use by and large, the current laws of physics to answer what a ‘true’ Nothingness could be and what properties it might have. The most prominent is the M-theory based on Supersymmetry that supports the so-called multi-verse theory (the existence of infinite number of universes) and the only theory (still not fully confirmed) that links gravity with two fundamental forces of nature: bosons and fermions.

For our needs it is enough to assume that whichever theory we use the only important issue is that most likely a “true” Nothingness does not exist. Nothingness has a property, and the current theories only need to prove that a given description of the properties of Nothingness is plausible. From a human perspective, Nothingness is the opposite of Consciousness because only if we are conscious can we be aware of our existence and the Existence, Non-Existence, or Non-Awareness

The emergence of ‘Something’ out of ‘Nothing’ is only meant to give you a better appreciation of what a momentous event the birth of Superintelligence would be and how it could eventually evolve. The first stage gives a snapshot of how it all has begun, starting with just a quantum field, which pre-existed the birth of our Universe, and culminating with today’s planet Earth. In stage two, I present my schedule for delivery of a mature Superintelligence. Finally, in stage three, I will present my personal view, based on dozens of scenarios published by particle physicists, astronomers, and futurists, on how the Universe with Superintelligence might evolve.

Stage 1: The evolutionary perspective on the emergence of human species

1. Nothingness 'has' no beginning and no end – it is eternal
2. Even Nothingness, as anything else, must have a property. It is believed that the property of Nothingness is a quantum field – “pure” energy
3. “Pure” energy manifests itself as Strings according to Supersymmetry or M-theory. Strings are one dimensional ‘packets’ of energy varying in size from below the Planck's scale (10^{-35m}) to the size of the Universe. They have charge, vibration frequency, gravitational force (gravitons) and mass, if crossing electromagnetic field or Higgs' field
4. Quantum field obeys quantum mechanics laws, among others the Heisenberg's Uncertainty Principle, which can sometimes lead to spontaneous imperfections in the field
5. Those imperfections appear “like waves on an ideally flat ocean’s surface”

6. Like any waves, quantum field waves have troughs and crests, which are positive and negative energies
7. On average the sum of positive and negative energies is 0
8. So, Nothingness is not really “nothingness” as we understand it. Sometimes, when imperfections occur, it becomes “Something”
9. That’s why I believe over Eternity, Nothingness is “true” Nothingness **on average**
10. Following the Heisenberg's Uncertainty Principle, the fluctuations in energy levels may lead to spontaneous creations of Big Bangs, creating material Universes
11. According to Multiverse Theory there could be infinite number of Universes emerging from those fluctuations, when the conditions are just right for converting energy into mass and creating a material world. Only some of them are stable enough, having the right constants (e.g. proportion of matter and antimatter, initial temperature, and gravity)
12. Our Universe was also started that way about 13.8bn years ago. It began with a period of so-called Inflation lasting between 10^{-35s} and 10^{-32s} when it expanded from zero size to the size of a grapefruit, ensuring the uniformity of the Universe
13. Since then our Universe has been constantly expanding leading to the creation of first stars after about 100 million years (the discovery made in 2018)
14. About 5bn years ago our Sun was created
15. 4.5bn years ago our Planet was formed
16. 3.5bn years ago life began on earth
17. A few million years ago man evolved from the animal kingdom
18. About 10,000 years ago our civilization begun
19. About 10 years ago, first artificial intelligence machine (IBM's Watson) beat humans in American Jeopardy game.
20. In 2016 Alpha-Go beat the Grand Master in the Go-Go game. It was built on the hypothesis that our mind organizes knowledge/intelligence in several layers. That discovery has sparked off almost faster than exponential progress in AI.

Stage 2: The Pathway to Superintelligence

So, this is where we are today. Below are my predictions on AI development and other key innovations that will ultimately lead to the emergence of Superintelligence:

2024 – A primitive cognitive AI Agent is aware of where it is, what it is doing, what it must achieve and what’s its relationship with its ‘master’ (it understands this is the man who makes ultimate decisions). It passes an advanced Turing test

2025 – The first Transhuman with a fully thought-controlled brain implant can do some Google searches by thought alone and storing and later on retrieving that information by accessing a private external memory and a processing unit

2026 – An AI agent with a wireless access to the fastest Supercomputer exceeds the intelligence of any human but in a single domain only. It can make hours-long unscripted conversations with any human on a specific subject area. It can also increase its understanding of other knowledge domains by self-learning or in conversations

2027– A Transhumans’ team has been created that can communicate wirelessly using brain implants

2028 – Approved Transhumans control the development of Superintelligence

2030 - Approved Transhumans are elected to political bodies, such as the European Federation

2033 – An AI agent is created, which exceeds human intelligence in many domains and can discuss for hours on any subject with any human. It has a cognition level of a teenager. It can be unpredictable but is still controlled by Transhumans.

2035 – Transhumans can read each other’s mind (in a broad sense), and if allowed, can also read any ordinary human mind (only knowing the subject area with no details). They can also control, if permitted, other humans’ subconsciousness, by creating preferences in their minds (similar to hypnosis but could in principle be done against the person’s will)

2042 – The first version of a mature Superintelligence is created, completely controlled by the authorized Transhumans. It has full cognition of an adult but may not be conscious yet (perhaps does not have to be in a human sense)

2045 – A fully mature Superintelligence is born, but which is willingly executing any requests from authorized humans. It is controlled by Transhumans, some of whom were elected as the leaders of the Human Federation

2047 – A Technological Singularity is born, fully controlled by Transhumans. Superintelligence improves its performance and intelligence almost exponentially

2050 – Superintelligence rules Humanity in every aspect of human life. The world of abundance has been created. Nobody has to work. However, most people have severe mental problems caused mainly by their inability to adapt to the pace of change.

What happens next is a pure speculation. This is my ‘gut’ feeling:

Stage 3 – Posthumans living as digital beings

2060 – First successful mind of a Transhuman is upload to Superintelligence. A Posthuman species has been born

2070 – Posthumans retain a complete control over Superintelligence

2200 – Most biological humans have become Posthumans. The twilight of Homo Sapiens is nearly there

From today's perspective, what is important, is to prepare for a selection of choices around 2040 – 2100, about how we want our species to evolve. This is of course a big assumption that we will survive unscathed as a species, with no existential risk wiping us out, that the Superintelligence will emerge as human friendly, and that we shall have enough control over it to make such a momentous decision. But there is still one more condition on which our choice will depend – can Superintelligence become conscious. What would be the consequences of developing Superintelligence, which in principle (confirmed by future discoveries) could never be conscious, because consciousness requires a biological substrate?

The events from 2045 onwards will start to evolve very rapidly. Very soon after it has been created, Superintelligence, embodied in whatever form, will be capable of redesigning itself. Repetitions of this cycle will result in the so called runaway effect, or an intelligence explosion, where such a smart machine designs successive generations of increasingly improved, powerful copies of itself, creating intelligence far exceeding human intellectual capacity. Once Superintelligence has reached that capability it may be impossible for a human to comprehend and even less so, to control it. Humanity will have reached the point known as **Technological Singularity**. Ray Kurzweil predicts that Technological Singularity event will happen by 2045.

If we manage to turn Superintelligence into our friend and let it arrive at the Technological Singularity event, assuming we would still have control over its evolution (e.g. via linking its goals to our most important human values), then the question is what will be our choice. For example, we can let it evolve itself and, so to speak, let it fly off and leave us alone. That may be possible. However, it is unlikely we will choose this option, since as someone said, 'one cannot uninvent the atomic bomb'. We are inquisitive and innovative beings and it is difficult to imagine we would like to return to 'less developed' world if it was at all possible.

We have to assume that once the Technological Singularity is there, it will become our Master by default, even if our values will have been embedded in its overall decision-making pattern. Its knowledge, choices of important decisions for humanity and overall comprehension of the world around us and the Universe in general, will be unimaginably greater than our own capabilities.

So, we shall have two options (assuming there are both possible). The first one is to upload (copy) our personality, memories, and consciousness by reading our

brain (although that may not be enough) onto a digital platform (a chip embedded in some material resembling our bodies). The second option is to merge our minds in a digital form with Superintelligence. From then on it is a pure speculation about how Humanity might evolve into the long, long future. However, in the next few decades we may be forced to make that biggest decision in the history of Humanity on **how** we want to evolve as a species. Therefore, let me take you on my own journey into the unknown because the conclusions may be useful for making such a decision.

Chapter 2

The risk arising from a hostile Superintelligence

I would now like to concentrate on the risk for Humanity arising from the advent of Superintelligence. The key aspect of the assessment of that risk for Humanity is to ask a basic question about key values that define Humanity, what is good and what is right. I will develop this argument further on. So, paradoxically, Superintelligence forces us to answer these questions more meaningfully than ever before.

In my view, Superintelligence represents the most dangerous risk. Why? Because it is almost certain to happen, if we do nothing, unlike natural pandemics that may not happen at all, even if we do not apply any counter measures, since it is a lottery type risk. The second reason why it is so dangerous is that it may happen much earlier than the risk mostly talked about in recent years – the climatic catastrophe. The risk coming from Superintelligence is more likely to happen in the next 50 years rather than next century. On the other hand, I believe that if we manage to produce the so called “friendly” Superintelligence, then instead of becoming the biggest risk, it itself can help us reduce other anthropogenic risks, such as climate change.

The Superintelligence is defined as a type of artificial intelligence that would surpass even the smartest humans. The threat stems, from even the slightest misalignment of our values and Superintelligence’s objectives, or its “values”. If this happens, even when the corresponding goals appear benign, it could be disastrous. Nick Bostrom quotes a scaring example that involves a Superintelligence programmed to “maximize” the abundance of some objects, like paperclips. This could lead Superintelligence to harvest all available atoms, including those in human bodies, thereby destroying humanity (and perhaps the entire biosphere⁽⁷⁷⁾). In addition, there are multiple ways that Superintelligence could become outright malevolent toward humanity, as University of Louisville computer scientist Roman Yampolskiy outlines in a recent paper (78).

- Preventing humans from using resources such as money, land, water, rare elements, organic matter, internet service or computer hardware;
- Subverting the functions of local and federal governments, international corporations, professional societies, and charitable organizations to pursue its own ends, rather than their human-designed purposes;
- Constructing a total surveillance state (or exploitation of an existing one), reducing any notion of privacy to zero – including privacy of thought;

- Enslaving humankind, restricting our freedom to move or otherwise choose what to do with our bodies and minds, as through forced cryonics or concentration camps;
- Abusing and torturing humankind with perfect insight into our physiology to maximize amount of physical or emotional pain, perhaps combining it with a simulated model of us to make the process infinitely long.

It would be impossible to provide a complete list of negative outcomes of an AI agent with general reasoning ability that he would be able to inflict. We can expect a lot of these sorts of attacks in the future. The situation is even more complicated once we consider systems that exceed human capacity. Superintelligence may be capable of inventing dangers we are not even capable of predicting or imagining

So, we have to accept that the creation of Superintelligence poses perhaps the most difficult long-term risks to the future of Humanity. Phil Torres identifies several issues here, saying that the first one is the amity-enmity problem: the AI could dislike us for whatever reason, and therefore try to kill us. The second risk is the indifference problem: the AI could simply not care about our well-being, and thus destroy us because we happen to be in the way. And finally, there is yet another problem, which he calls “the clumsy fingers problem”: the AI could inadvertently nudge us over the cliff of extinction rather than intentionally pushing us. This possibility is based on the assumptions, which states that higher levels of intelligence aren’t necessarily correlated with the avoidance of certain kinds of mistakes. He warns that the fruits of our ingenuity — namely, dual-use technologies — have introduced brand new existential risk scenarios never before encountered by Earth-originating life. Given the immense power of Superintelligence, e.g. it could manipulate matter in ways that appear to us as pure magic, it would be enough to make a single error for such a being to trip humanity into the eternal grave of extinction⁽⁷⁹⁾.

The existential risk posed by Superintelligence does not depend on how soon one is created; it merely concerns us what happens once this occurs. Nonetheless, a survey of 170 artificial intelligence experts made in 2014 by Anatolia College philosopher Vincent C. Müller and Nick Bostrom suggests that Superintelligence could be on the horizon. The median date at which respondents gave a 50 percent chance of human-level artificial intelligence was 2040, and the median date at which they gave a 90 percent probability was 2075. This prediction is further away than 2045 given by Ray Kurzweil, that I quoted before. In any case, if they are correct, some people around today will live to see the first Superintelligence—which, as British mathematician I. J. Good observed in 1966, may be our last invention.⁽⁸⁰⁾

Physicist Stephen Hawking, Microsoft founder Bill Gates and SpaceX founder Elon Musk expressed concerns about the possibility that AI could evolve to the point that humans could not control it, with Hawking theorizing that this could "spell the end of the human race". In 2009, AI experts attended a conference hosted by the Association for the Advancement of Artificial Intelligence (AAAI) to discuss whether computers and robots might be able to acquire any sort of autonomy, and how much these abilities might pose a threat or hazard. They noted that some robots have acquired various forms of semi-autonomy, including being able to find power sources on their own and being able to independently choose targets to attack with weapons. They also noted that some computer viruses can evade elimination and have achieved "cockroach intelligence." They concluded that self-awareness as depicted in science-fiction is probably unlikely, but that there were other potential hazards and pitfalls. Various media sources and scientific groups have noted separate trends in differing areas, which might together result in greater robotic functionalities and autonomy, and which pose some inherent concerns. One of those well-known AI experts, Eliezer Yudkowsky, believes that risks from AI are harder to predict than any other known risks. He also argues that research into AI is biased by anthropomorphism. He claims that since people base their judgments of AI on their own experience, they underestimate its potential power. He distinguishes between risks due to technical failure of AI, which means that flawed algorithms prevent the AI from carrying out its intended goals, and philosophical failure, which means that the AI is programmed to realize a flawed ideology⁽³⁶⁾.

The reason why I believe Superintelligence is the biggest risk is that it is the one that may arrive in an inferior, "half-baked" form. I called it an Immature Superintelligence, which is most likely to appear by about 2030. There is certainly no need for Superintelligence to be conscious to annihilate Humanity. It is worth to remember what kind of panic and material loss was caused by the 'WannaCry' ransom virus, on 13th May 2017, believed to have been stolen from the US National Security Agency, almost infinitely primitive by comparison with Superintelligence. The virus was reportedly spread out by North Korea. As reported by BBC, it targeted computers running the Microsoft Windows operating system by encrypting data and demanding ransom payments in the Bitcoin cryptocurrency. It was reported that within a day it had infected more than 230,000 computers in over 150 countries, including Russia and China. Parts of the United Kingdom's National Health Service were infected, causing it to run some services on an emergency-only basis during the attack. Spain's Telefónica, FedEx and Deutsche Bahn were hit, along with many other countries such as Russia, the Ukraine and Taiwan. Only by sheer coincidence the attack was stopped within a few days by Marcus Hutchins, a 22-year-old web security researcher, who discovered an effective solution⁽⁸¹⁾.

This example shows that it is enough for AI agent to be more intelligent in a specific area than any human and that its intelligence being digital can increase exponentially. It would be enough for such an entity to annihilate Humanity if for example it had slightly misaligned objectives or values with those that we share. Such misalignment may then lead immediately to the point of no-return, by triggering the so-called run-away scenario of Technological Singularity. Malhar Mali in his interview with Phil Thores of X-Risks, puts it very clearly:

“When it comes to creating the Superintelligence, the coding becomes important. Because there's a difference between ‘do what I say’ and ‘do what I intend.’ Humans have this background knowledge that enables us to figure out what people actually say - in a context-appropriate way. But for an A.I., this is more of a challenge... it could end up doing exactly what we say but in a way that destroys the human race.”⁽⁸²⁾

This kind of risk is well illustrated by the Greek legend about Tithonus, the son of Laomedon, the king of Troy. When Eos (Aurora), the Goddess of Dawn, fell in love with Tithonus, she asked Zeus to grant Tithonus eternal life. Zeus consented. However, Eos forgot to ask Zeus to also grant him eternal youth, so her husband grew old and gradually withered.

It is difficult to imagine at first what kind of damage a wrongly designed Superintelligence can do. The most dangerous period for Humanity, which in my view will last for about one generation, has already started. If we somehow survive, by managing the damage that will be occurring from time to time, and maintain our control over Superintelligence, it will be Superintelligence itself that will help us to minimize other risks.

If one minor computer virus such as WannaCry, quoted earlier, can do such damage then imagine what might be expected even from a relatively primitive Superintelligence if it is applied in a full-scale cyberwarfare. In theory, such a cyberwarfare could trigger off a cascade of a series of non-existential risks, which could combine into an existential risk. For example, what would be the consequences if North Korea, or a super-rich derailed billionaire acquire the capability of cracking any password within minutes, using quantum-computing (China probably already has this capability). It could then also get access to the most important state and military secrets, including access to firing nuclear weapons.

The good news is, that at the same time as it would be possible to crack any password using quantum computing technology it would also be possible to protect access to various state guarded secrets by applying quantum encryption. It has already been proved to work by China in February 2018, when not only the passwords but also the whole content (a video) was quantum encrypted. Since quantum encryption makes it physically impossible to get access to

protected information by cracking a password, this will reduce the risk of a full-scale cyber war.

Viruses are only one aspect of the damage that even basic IT can do. There have been many other IT-generated non-virus-related damages. More recently, we are seeing the first occurrences of the damage done by the so-called narrowly focused AI systems. These are AI agents that excel in only one or two domains. The damage done by today's AI systems included market crashes, accidents caused by self-driving cars, intelligent trading software, or personal digital assistants such as Amazon Echo or Google Home.

Elon Musk, the founder of Tesla, Space X and the Neuralink, a venture to merge the human brain with AI, has been urging governments to take steps to regulate the technology before it's too late. At the bipartisan National Governors Association in Rhode Island in July 2017 he said: "AI is a fundamental existential risk for human civilization, and I don't think people fully appreciate that." He also said, he had access to cutting-edge AI technology, and that based on what he had seen, AI is the scariest problem. Musk told the governors that AI calls for precautionary, proactive government intervention: "I think by the time we are reactive in AI regulation, it's too late," he said. ⁽⁸³⁾

Most people still think that AI should continue being developed like all previous technologies. But as with each technology, the more advanced AI becomes the more people it can affect. Even AI researchers still behave and develop their AI agents as they were a similar piece of technology, as a rudimentary IT program. Even if we assume this argument, many human inventions have potentially both a positive and a negative effect. Suffice to give two examples; nuclear energy and the Internet. Although it is true that in principle AI is a tool (so far) as any other invention before, i.e. not inherently good or bad, it differs from all previous inventions in that it **could** lead to unimaginable unintended consequences. So, how could we minimize the risk arising from Superintelligence? That's why the next section is about.

Chapter 3

Minimizing the risks of a hostile Superintelligence

Setting the rules for controlling Superintelligence

I would now like to focus on how we can minimize the risks that the advent of Superintelligence may create for Humanity. We have to make absolutely sure that however we approach the development of Superintelligence, we have covered all conceivable risks resulting from its activity. Otherwise, we may deliver the agent that will annihilate Humanity.

The good news is that there are already some countermeasures in place, which aim at minimizing the risk of Superintelligence deployment. Until 2016 AI development was broadly guided by Three Laws of Robotics described by the science fiction writer Isaac Asimov in 1942 in a short story "Runaround" and later on repeated in his 1950 book "I Robot". They are:

- A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws ⁽⁸⁴⁾.

These principles have now been replaced by 23 Asilomar Principles agreed at the Beneficial AI Conference at Asilomar, California on 5th January 2017 and signed by over 2,000 AI experts in the first three months. It is intended to be constantly evolving as new AI challenges appear. They have been split in three areas:

Research issues

1. Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence
2. Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:
 - How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?

- How can we grow our prosperity through automation while maintaining people's resources and purpose?
 - How can we update our legal systems to be fairer and more efficient, to keep pace with AI, and to manage the risks associated with AI?
 - What set of values should AI be aligned with, and what legal and ethical status should it have?
3. Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policymakers
 4. Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI
 5. Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards

Ethics and Values

6. Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible
7. Failure Transparency: If an AI system causes harm, it should be possible to ascertain why
8. Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority
9. Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications
10. Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviours can be assured to align with human values throughout their operation.
11. Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
12. Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyse and utilize that data.
13. Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
14. Shared Benefit: AI technologies should benefit and empower as many people as possible.
15. Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
16. Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

17. Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.
18. AI Arms Race: An arms race in lethal autonomous weapons should be avoided.

Longer-term Issues

19. Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.
20. Importance: Advanced AI could represent a profound change in the history of life on Earth and should be planned for and managed with commensurate care and resources.
21. Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.
22. Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.
23. Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.

How to tame Superintelligence?

Stephen Hawking, the renowned physicist, and who was one of the most alarmed people among the scientists regarding the risks posed by Superintelligence said that “if Superintelligence isn’t the best thing to ever happen to us, it will probably be the worst”.

That’s why people like Nick Bostrom, one of the top experts on Superintelligence, think we need to invent some controlling methods to minimize the risk of Artificial General Intelligence (AGI) going terribly wrong. He defines these methods in his book “Superintelligence” ⁽⁷⁷⁾. For our purpose I will try to provide a layman’s description of what it really means and what are the consequences for controlling the risks emerging from Superintelligence. The most important point is that these controlling methods must be in place **before** Superintelligence arrives, i.e. latest in the next decade.

Nick Bostrom identifies the ‘control problem’ as the ‘principal-agent’ problem, a well-known subject in economic and regulatory theory. The problem can be looked from two perspectives:

- **The first ‘principal-agent’ problem:** e.g. the problem faced by a client wanting to buy a house and employing an estate agent to fulfil exactly *his (client’s)* needs. In this scenario, the client is the *principal* (the person who wants some task to be performed in accordance with his interests), and an estate agent is the *agent* (the person carrying out the tasks on my behalf).
- **The second ‘principal-agent’ problem:** e.g. the problem where the estate agent thinks primarily about *his own* interest e.g. to get the best possible agent’s fee

He dedicates a whole chapter to identify potential solutions. Since the publication of the book in 2013, they have been widely discussed in the AI community on how to turn them into practical tools.

The ‘Control Problem’ involves human principals (sponsors or financing institutions) and human agents (AI developers). At some stage there will be an AI project to develop Superintelligence (AGI). It may be launched by one of the big IT/AI companies such as Google, Microsoft, IBM, or Amazon. But it is also quite likely it will be initiated by some wealthy AI backers, which is already happening. Probably the most prominent among such people deeply involved in various top AI initiatives is Elon Musk. He is the founder of PayPal – a credit transaction payment system, SpaceX – rocket company, Hyperloop – a network of underground trains travelling at speeds of nearly 1,000 km/h, Neuralink a brain-computer interface venture, and several other large-scale initiatives such as sending 1 million people to Mars by 2050. The second one is Jeff Bezos, the founder of Amazon and the richest man on the planet with assets of about \$100bn who is deeply involved in AI. His micro AI-product called Alexa Echo was sold to over 20m people by the end of 2017.

Such sponsors will need to ensure that AI developers carry out the project in accordance with their needs. They would also want to ascertain that they **understand** their sponsors’ needs correctly **and** that the developed AI product, which may turn into Superintelligence will also understand and obey humans as expected. Failure to address this problem could become an existential risk for Humanity.

Capability Control Method

Bostrom specifies four possible solutions, which he calls the “**Capability Control Method**”. Its purpose is to tune the capabilities of superintelligent agent to the requirements of humans in such a way that we stay safe and have the ultimate control of what Superintelligence can do.

Boxing Methods of Control

This is perhaps the simplest and most intuitively compelling method of controlling Superintelligence - putting it into a metaphorical “box” i.e. a set of protocols that constrain the way, in which Superintelligence could interact with the world, always under the control of humans. It is often proposed that as long as Superintelligence is physically isolated and restricted, or “boxed”, it will be harmless.

A typical Superintelligence will be a superbly advanced computer with sophisticated algorithms (procedures how to process information) and will have three components: a sensor (or input channel); a processor; and an actuator (or output channel). Such superintelligent agent will receive inputs from the external world via its sensors e.g. Wi-Fi, radio communication, chemical compounds, etc. It will then process those inputs using its processor (computer) and will then respond (output information or perform some action using its actuators. An example of such action could be very mundane such as advising on which decision should be made, switching on or off certain engines, or completing financial transactions. But they could also be potentially significant e.g. whether a chemical compound would be safe for humans at a given dose.

However, it is unlikely that a superintelligent agent could be boxed in this way in the long term. Once it becomes superintelligent, it could persuade someone (the human liaison, most likely) to free it from its box and thus it would be out of human control

Incentive Methods

Bostrom refers to the second capability control method as the “incentive” method. The idea is that if you create the right “incentive environment” Superintelligence wouldn’t be able to act in an existentially threatening manner. This is in some way an analogy to how to bring up a child. A child has its own goals, which may not be good for itself or the people around it right now or in the future. So, a good teacher can motivate his child in such a way that it behaves in morally and socially acceptable ways.

Stunting

“Stunting”, as the name implies, involves hampering or disabling Superintelligence in some way. A good example would be running Superintelligence on a slow hardware, reducing its memory capacity, or limiting the kind of data it can process. Bostrom argues that the use of stunting poses a dilemma. Either we stunt Superintelligence too much and it just becomes another

“dumb” piece of software; or we stunt it too little and it would be capable of overcoming its disabilities. Getting the balance just right could be pretty tricky.

Tripwiring

This is a different method of capability control. It involves building into any AI development project a set of “tripwires” which, if crossed, will lead to the project being shut down and destroyed. Bostrom identifies three types of tripwire:

- **Behavioural Tripwires:** A network of detectors would be placed around a “boxed” AI in order to detect any attempt to breach the containment, e.g. any attempt to make radio communication.
- **Ability Tripwires:** The Superintelligence would be regularly tested for its capabilities. If at any point new abilities are detected or if it gets close to some capability threshold, it would be shut down.
- **Content Tripwires:** The Superintelligence’s internal processing could be constantly monitored for any suspicious changes, e.g. changes in how it represents its final values.

Bostrom thinks that tripwires could be useful, particularly during the development phase if used in conjunction with other methods. But, unsurprisingly, he also thinks that they too have shortcomings. An advanced Superintelligence could subvert a tripwire in an unexpected way. He also notes that project developers working on Superintelligence could grow impatient if tripwires repeatedly hamper their progress. They might undermine any safety advantage gained by the tripwire system.

Motivation Selection Method

The second set of tools to control the undesired behaviour of Superintelligence is to try to motivate it to pursue the goals that are in our (human) interest, and that is why this approach is called the “**Motivation Selection Method**”. John Danaher provides a summary of these methods in his article “Bostrom on Superintelligence: Limiting an AI's Capabilities”⁽⁸⁵⁾, parts of which I have used to convey below the essence of Motivation Selection in a less technical way.

It is in some way an extension of the ‘Incentive Method’ from the Capability Control set of tools. Bostrom is clear as with the Control Problem approach, that this set of methods would have to be implemented **before** an AI achieves Superintelligence. Otherwise, the Superintelligence could have a decisive strategic advantage over human beings, and it may be impossible to constrain or limit it in any way.

That is why I have already stressed that **we have really about one decade, till about 2030, to implement mechanisms of controlling Superintelligence.**

Direct Specification

This involves programming the Superintelligence directly with the “right” set of motivations. What could go wrong if a robot always follows Asimov’s first law which I mentioned earlier? Of course, anyone who has read the book will know that lots can go wrong. Laws and rules of this sort are vague. In specific contexts they could be applied in very odd ways, especially if the robot has a very logical or literalistic mind. Take the first law as an example. It says that a robot may not, through inaction, allow any human to come to harm. This implies that the robot must at all times be seeking to avoid possible ways, in which humans could come to harm. A superintelligent robot, with a decisive advantage over human beings, might decide that the safest thing to do would be to put all humans into artificially induced comas. It wouldn’t be great for them, but it would prevent them from coming to harm.

So, anyone who has studied the development and application of human laws will be familiar with this problem. The drafters of those laws can never fully anticipate every possible future application. The same will be true for AI programmers and coders.

Domesticity

The second suggested method of motivation selection is called “domesticity”. The analogy here might be with the domestication of wild animals. Dogs and cats have been successfully domesticated and tamed from wild animals over many generations. The suggestion is that something similar could be done with superintelligent agents. They could be domesticated. The classic example of a domesticated superintelligence would be the so-called “oracle” device. This functions as a simple question-answering system. Its final goal is to produce correct answers to any questions it is asked. Even a simplistic micro AI gadget like “Alexa”, which I mentioned earlier, can already do that. Superintelligent agents would usually do just that from within a confined environment (a “box”). This would make it domesticated, in a sense since it would be happy to work in a constrained way within a confined environment.

However, giving Superintelligence the seemingly benign goal of giving correct answers to questions could have startling implications. To answer the question, Superintelligence may require quite a lot of information, as anyone that has tried to talk with Google Home or Amazon Alexa appreciates. Once that information is stored in its memory, it will make the superintelligent agent more

knowledgeable and more capable, increasing the risk of its misbehaviour, including a potential ‘runaway’, i.e. a total loss of control by humans.

Indirect Normativity

The third possible method of motivation selection is Indirect Normativity. The idea here is that instead of directly programming ethical or moral standards into Superintelligence, you give it some procedure for determining its own ethical and moral standards. If you get the procedure just right, Superintelligence might turn out to be benevolent and perhaps even supportive of human interests and needs. Superintelligence is to function much like an ideal, hyper-rational human being, which can “achieve that which we would have wished it to achieve if we had thought about the matter long and hard” (Bostrom “Superintelligence” p. 141) (77)

One of the problems with this method of motivation selection is ensuring you’ve got the right norm-picking procedure. Getting it slightly wrong could have devastating implications, particularly if a superintelligent machine has a decisive strategic advantage over us.

Mind Augmentation

This is quite different from the methods discussed thus far. There, the assumption was that Superintelligence would be delivered from scratch through a series of ever more intelligent AI agents. This assumes that we start with a system that has the “right” motivations and we increase its intelligence from there. The obvious candidate for such a system would be a human being (or a group of human beings). We could simply take their brains, with their evolved and learned motivations, and augment their capabilities until we reach a point of Superintelligence. (Ignore, for now, the ethics of doing this.) Such an approach is favoured by Transhumanists and Posthumanists, who envisage that at some stage human species will merge with Superintelligence.

As Bostrom notes, augmentation might look pretty attractive if all other methods turn out to be too difficult to implement. Furthermore, it might end up being a “forced choice”. If augmentation is the only route to Superintelligence, then augmentation is, by default, the only available method of motivation selection. Otherwise, if the route to Superintelligence is via the development of AI, augmentation is not on the cards.

But a “solution” to the control problem by augmentation is not perfect either. If the system we augment has some inherent biases or flaws, we may simply end up exaggerating those flaws through a series of augments. It might be wonderful to augment a Florence Nightingale to Superintelligence, but it might be

nightmarish to do the same with a Hitler. Furthermore, even if the starter-system is benevolent and non-threatening, the process of augmentation could have a corrupting effect.

Uploading the Values of Humanity to Superintelligence

It is clear from the last paragraph in the preceding section that the least risky strategy, for delivering Superintelligence would be the process of mind augmentation (although it may also have some inherent danger). Apart from enormous technical problems that will emerge, **the equally important issue will be the kind of values the new augmented species should have**, which would become more than just a digital Superintelligence. That's why the need to define top values of Humanity, the foundation of human ethics, is so important (we shall discuss it in Part 3).

These values will constitute the new Human Values Charter. We will then need to establish certain procedures, perhaps enshrined in laws regarding the transfer of these values into various shapes and types of AI robots and humanoids. This would create a kind of a framework where the Human Values Charter becomes the core of every AI agent's 'brain'. Such a framework would be a boundary beyond which no AI agent could act and implemented using certain guidelines, such as 23 Asilomar principles mentioned earlier. Only then could the developers define specific goals and targets for such AI agents always referencing the Human Values Charter as constraints to agents' objectives. In practical terms the best way forward could be to embed the Human Values Charter into a sealed chip that cannot be tampered with, perhaps using quantum encryption, and implant it into any intelligent AI agent.

I believe the way Superintelligence behaves and how it treats us will largely depend on whether at the Singularity point it will have at least basic consciousness. My own feeling is that if a digital consciousness is at all possible, it may arrive later than the Singularity event (*this subject is explored in more detail in my previous book "Who could save Humanity from Superintelligence?"* ⁽⁸⁶⁾). In such case, one of the mitigating solutions might be, assuming all the time that Superintelligence will from the very beginning act benevolently on behalf of Humanity, that decisions it would propose would include an element of uncertainty by taking into account some emotional and value related aspects. We will develop the question of values later on in a broader context.

In the long-term, I think there is a high probability that the human race as we know it will disappear. Why should we be the only species not to become extinct? After all, everything in the universe is subject to the law of evolution.

We have evolved from apes and we will evolve into a new species unless some existential risks will annihilate civilization before then. We can speculate whether there will be augmented humans, synthetic humans, or entirely new humanoids, i.e. mainly digital humans with uploaded human minds or even something entirely different that we cannot yet envisage. It is quite likely, that humans will co-exist with two or even three species for some time but ultimately, we humans will disappear at some stage.

The next question is should we “allow” the new breed of humanoids to define ethics for themselves or should they be jump-started by our ethics. In my view, we should try as much as possible make a transfer of human ethics into the new species. Therefore, whichever organisation takes over the task of saving Humanity, there is an urgent need to formally agree the renewed set of Universal Values and Universal Rights of Humanity so that the world could reduce the level of existential risks before the AI-based humanoids adopt them. Beyond that, however, when our human ethics is re-defined at some stage by AI-based humanoids, the values with the corresponding ethics are bound to change. Ethics is not static.

Chapter 4

Can Superintelligence help reduce existential risks?

If we design Superintelligence properly, its emergence does not have to be a catastrophic event in the human history. In such case a new period in Humanity's existence will start – the Big Coexistence. A friendly Superintelligence, which genuinely cares about our well-being, could guide us through various hazards that endanger human species. This covers both anthropogenic and non-anthropogenic risks, such as asteroid impact, which if detected early could be put on a trajectory bypassing Earth. More importantly, Superintelligence may have different ways of analysing potential risks, based on different idea-generating mechanisms, of which we humans could be totally unaware.

Once the Big Coexistence starts the human species may have very little influence on its own future. Some people with extended AI capabilities will become humanoids, some will decide to merge with Superintelligence and perhaps keep living in a digital form. Delivering best human ethics in the form of widely agreed Universal Human Values and Rights may be the biggest legacy that Humanity will deliver to the new species of intelligent and conscious beings. After that the next generation of “ethical” Superintelligence may itself redefine ethics of the kind we cannot even imagine.

Assuming Superintelligence develops its capabilities gradually and would be under our full control, becoming quite possibly a conscious Superintelligence, the question is how it could directly help in mitigating all other existential risks. Superintelligence, if properly designed and managed can deliver incredible benefits to Humanity and at the same time make our future much safer. This is not yet an overwhelming view among the scientists. A lot depends on how we prepare ourselves for this moment and whether there will be any intervening catastrophic events, which would bury the dream of Superintelligence and possibly be the end of civilization (e.g. engineered, untreatable pandemics). Among the optimists we have Max Tegmark, a well-known cosmologist. In his book “Life 3.0: Being Human in the Age of Artificial Intelligence” he gives quite an optimistic view on what Superintelligence can do for us ⁽⁸⁶⁾.

In an interview with Clive Cookson ⁽⁸⁷⁾, Tegmark remains convinced that barring some cataclysmic disaster in the next few decades, Superintelligence will take over the world. But he believes that we can shape the way this happens, including embodying human values. In his view, the next few decades on Earth could have cosmic significance, determining “nothing short of the ultimate future of life in our universe”. Given that our galaxy has about 100bn planets and there are 200bn

galaxies in the universe, most astronomers maintain that extra-terrestrial intelligence must be widespread. Since Superintelligence is almost inevitable, we should make every effort now to ensure that it is friendly.

There are a number of computer scientists who believe Superintelligence will be in a form of human-machine hybrids such as cyborgs with the uploaded human brains into computer intelligence. However, Tegmark disagrees with them. Clean-slate Superintelligence will be much easier to build and, even if cyborgs and uploads are introduced, their human component is likely to make them uncompetitive in the long run against pure Superintelligence. Once it has exceeded human abilities, our knowledge of physics suggests that it will advance rapidly beyond the point that biological intelligence has reached through random evolutionary progress. As Tegmark points out, “information can take on a life of its own, independent of its physical substrate”. In other words, any aspect of intelligence — presumably including consciousness — that evolved in flesh, blood and carbon atoms can exist in silicon or any other material. No one knows what the next blockbuster substrate will be but Tegmark is confident that the doubling of computing power every couple of years will continue for a long time. (I might agree with this view with one proviso. Transhumanism, i.e. blending part of our body with Superintelligence should be seen as a transitory phase to a fully digital form. That would be a much safer passage and would give Transhumans more time to decide on the best way for the evolution of the new species.)

The fundamental limit imposed by the laws of physics of how fast computers can be is a billion, trillion, trillion times more powerful than today’s best computers. The intelligence explosion could propel AI across the universe, generating energy billions of times more efficiently than present-day technology. Tegmark describes candidate power sources such as black holes, quasars and a “sphalerizers” that convert heavy fundamental particles (quarks) into lighter ones (leptons). The message at the heart of Life 3.0 and Tegmark’s “beneficial AI” movement is that, since Superintelligence is almost inevitable, we should make every effort now to ensure that it emerges in a way that will be as friendly as possible to human beings, primed to deliver the cosmic inheritance we want. If we wait too long, it may be too late.

At present no one has a clear idea of how to achieve this. At a moral and political level, we need to discuss what goals and qualities to incorporate. At a technical and scientific level, researchers must figure out how to build our chosen human values into AI in a way that will preserve them after we have lost direct control of its development. Tegmark advances various options and scenarios in which Superintelligence plays the roles ranging from “gatekeeper” to “protector god”, “zookeeper” to “enslaved god”. “I view this conversation about the future of AI

as the most important one of our time. Life 3.0 might convince even those who believe that AI is overhyped, to join in” (87).

He is supported in such views by Stuart Russell, a British-American AI scientist. He proposes that to ensure that **the goal we have in mind will be correctly understood by Superintelligence, three principles must be observed**. I consider these principles probably the most practical solution that can actually work because it would make Superintelligence behave more like we do:

1. Superintelligence needs to know in minute detail, supported by thousands of examples, what are our top human values
2. Allow Superintelligence to have some margin of doubt both on the rationality of those values and then on their interpretation
3. Teach Superintelligence what these values really mean in practice by letting it observe for some time how people actually implement those values.

Assuming we teach Superintelligence, our values and have a full control of its activity, it can become an enormous help for the whole of humanity to solve almost any problem we have. All anthropogenic existential risks and even some risks stemming from natural disasters (such as super volcanos), including of course climate change, could be drastically lowered, or eliminated. The only caveat might be that for Superintelligence to help us successfully, we may need to trust its judgments and decisions and fulfil what is expected from us.

At this stage, my overall assumption is that we will somehow manage to control Superintelligence and make it our “best friend”. We should start developing practical measures right now by adopting 23 Asilomar Principles, defined earlier on, so that Superintelligence itself presents as low a risk to us as possible before it transforms itself into Superintelligence and becomes a Technological Singularity.

The next step would be to help Superintelligence to understand who we are as humans and what are our most important values. This is why it will be so critical to redefine our key human values on behalf of the whole Humanity because this will ultimately become a joint set of values shared by humans and Superintelligence. I discuss this in detail in Part 2 of the book.



3

PART 3
THE NEED FOR THE WORLD
GOVERNMENT

Chapter 1

Who could act as the future World Government?

Why do we need the World Government?

Anyone who wants to improve the situation and reduce some of the existential risks faces three problems:

- Existential risks require fast action, while the world's organisations act very slowly
- People want more freedom, while we need to sacrifice some of our freedoms and sovereignty for Humanity to survive
- Most people can't see beyond tomorrow and act emotionally, while we need to act rationally and see the long-term consequences of our actions.

Therefore, anybody that sees the need for the world to take an urgent action faces a difficult task when proposing pragmatic, fast and very radical changes to the ways the world is governed. It seems to me that the only realistic route for humans to take is to create a new organisation, which would have the capacity, resources and resolve to act on behalf of all of us in the hour of the emerging existential threat. To have any chance of successful delivery of its foremost objective, i.e. to protect Humanity against existential risks, such an organisation should have supranational powers exceeding any prerogatives of the existing international bodies, such as United Nations, NATO, or WTO. We need an organization that would resemble a World Government. Only such an organisation, which should be operational by about 2030, would have some chance of mitigating not just global political and social risks but all other existential risks mentioned before.

You may ask: Why do we need it so early, almost right now, what's so special about 2030? A simple answer is that two of the most significant existential risks may reach a tipping point by then.

The first one is Climate Change. Many climatologists believe that realistically, we must halt the average global temperature increase at 1.5C, and we are already quite close to that threshold. If we want to achieve the goal of a maximum global temperature of less than 1.5C then we must start dropping CO2 emissions well before 2030. If we continue emitting fossil fuels as before, continuously increasing CO2 level, then by 2030 the Earth's climate may reach a tipping point in many areas, leading to an uncontrolled global temperature increase.

The second and even more imminent threat is maturing AI. According to many eminent AI scientists, such as prof. Russel Stuart, we must have a full, global control over AI by 2030. Beyond that, even if we will be dealing with an Immature Superintelligence, it may already be too late to reign it in. Therefore, if we want to fight existential risks successfully, **we really have just about one decade to have a full, global control over all existential risks.**

A global control can only be effective if we have a global organization with stringent powers. The agenda of such an organisation should, therefore, be governed by one key issue: fighting existential risks. Any other objectives are subordinate to this goal, since if there is no Humanity and no civilisation there is no point to discuss other aims of such an organisation. This objective sets the agenda for the rest of the book.

As you shall see, I am deep down pragmatist and provide some solid arguments why a new organisation that might act as the World Government, co-ordinating all efforts to save Humanity from existential risks, cannot be created from scratch. The main reason is that we simply do not have enough time. If we, as humans, are serious about protecting our future, and indeed the species as such, the only alternative we have is to convert an existing organisation.

Having more sovereignty and more individual freedom, as the Brexit supporters in the UK dream about, while at the same time making our lives safer and more secure in this very fast changing world is rather unrealistic. You cannot have a cake and eat it. No nation can any longer be more secure by becoming more alienated and superficially more sovereign. Today, as never before in human history, our priority is to survive as a civilisation and a human species. That's how serious current existential risks are.

However, today we are still in a position to modify the course of events by making some fundamental changes in how Humanity's political and social order is set up and how we harness the technological progress for the benefit of the human race. That is what this book is mainly about - trying to answer the key question: how we can avoid the worse aspects of technological and biological advancement as well as disruptive tendencies in our societies. If we can do that in the course of the coming generation, then with the help of Superintelligence, we will create a wide range of benefits that will change our lives very positively beyond what most utopians were hoping for, delivering to us the world of abundance.

There are many scientists and experts in existential risks who call for immediate steps to be taken to form an organization that would act as the World Government as the only hope Humanity has in order to survive as a species. One of them was Stephen Hawking, the prominent theoretical physicist, who died in

March 2018, and who a year earlier said that: “Without the World Government technology will destroy us. This aggression may destroy us all by nuclear or biological war. We need to control this inherited instinct by our logic and reason”. In an interview with the Times he spoke about the dangers of Artificial Intelligence believing we need to establish a way of identifying threats quickly before they have a chance to escalate. He suggested that “some form of World Government could be ideal for the job, but it would itself create more problems and might become a tyranny...All this may sound a bit doom-laden, but I am an optimist. I think the human race will rise to meet these challenges.” (88)

That warning by Stephen Hawking about the World Government becoming a tyranny has to be taken very seriously indeed. That’s why I have put so much emphasis in this POSTHUMANS series about the necessary agreement on the new definition of Universal Values of Humanity that could become the foundation for a deep reform of democracy and the future Constitution of Humanity.

If we look at us, humans, from the perspective of International Space Stations (and many returning astronauts confirm that), it becomes very clear how unprepared we are to face existential risks. Here I would strongly agree with prof. Martin Rees, the former Astronomer Royal, who emphasizes it in his article “The world in 2050 and beyond” when he says that “Humanity is under long-term threat from anthropogenic global changes to climate and biodiversity – due to rising population, all more demanding of food, energy and other resources. All these issues are widely discussed. What’s depressing is the inaction – for politicians, the immediate trumps the long-term; the parochial trumps the global. **We need to ask whether nations need to give up more sovereignty to new organisations along the lines of IAEA, WHO, etc.**” (89)

The subject of the World Government is not new and there are a number of organisations that make specific proposals. One of them is the World Federalist Movement - Institute for Global Policy (EFM-IGP), founded in 1947 in the USA. It is a non-profit, non-partisan organization committed to the realization of global peace and justice through the development of democratic institutions and the application of international law. EFM-IGP includes a comprehensive set of programs that work to protect civilians from the threat of genocide, war crimes, and crimes against humanity; facilitate transparency in governance; increase access to justice; and promote the application of the rule of law. They work in partnership with the United Nations, governments, and other international and regional institutions, as well as with thousands of committed individuals around the world. ⁽⁹⁰⁾

Some of the aims of the World Federalist Movement are close to what I would see is necessary for Humanity to survive. The main difference is the key

objective. They hardly mention existential risks, or at least this is not their driving force, nor do they propose solutions, which in my view would be practical. The best evidence is that having been in existence for 70 years, they are nowhere near to achieving their goal. And this is certainly a great pity because we so badly need the World Government. The mistake that they make, in my view, is that they are idealistic in their objective (they want all nations to be governed by such an organization). But that is precisely why it is not achievable. However, their Manifesto is very useful as the first draft of the Humanity's Constitution since it goes much further and is more rigid than the UN Charter.

So, how popular would be this idea of the World Government? In May 2017, Global Challenges Foundation Global Risks Survey requested ComRes, a polling organisation, to make a Survey on global risks. They interviewed 1000 adults aged between 18 and 64 in each of these countries: Australia Brazil, China, Germany, India, South Africa, the UK, and the USA. Here are some of the results ⁽³⁵⁾:

- A majority (61%) of the general public in the eight countries surveyed consider the world to be more insecure today when it comes to global risks compared with two years ago. A quarter (27%), consider the world much more insecure today.
- Usage of weapons of mass destruction is ranked as the global risk needing the most urgent response (62%), followed by climate change (56%).
- Eight adults in ten (85%) think that the UN needs to be reformed to better address global risks. Only 7% do not think that it needs to be reformed.
- 71% of the general public across the eight countries think **that a new supranational organisation should be created** to make enforceable global decisions to address global risks.

Assuming that the sample of respondents reflects global population, the fact that nearly three quarters of the people across the globe support the creation of a supranational organisation, is very encouraging. In any case, we should be aware of the need of a long and fairly detailed campaign on existential risks similar, but more intensified, to the one promoting the necessity to combat climate change.

How realistic is this goal to set up such a new organization from scratch within the next 10 years? To answer this question, let's look at how long it took for some organizations to start their operation from the time of their inception. I have compiled the relevant data in the table below:

How long it took to establish the World's Political Organizations					
Organization	Date started	Declaration signed	Effect Date	Years	Comments
UDHR - Universal Declaration of Human Rights	1946	10.12.1948	1976	30	It was only in 1976 when more than 50 UN countries signed it making it legally binding. Former communist countries signed it in 1990'. Muslim countries declined to sign it, instead they signed the Cairo Declaration in 2000, compliant with Sharia law.
IPCC - Intergovernmental Panel on Climate Change	1988	14.061992	12.12.2016	26	RIO Declaration in 1992, Paris Agreement signed on 12.12.2015
UN - United Nations	12.06.1941	25.04.1945	24.10.1945	4	It was based on the League of Nations that existed for 20 years . That's why it only took 4 years. It does not have supranational powers because of articles 2 and 51 and that's why even Russia and China could sign it. All major powers signed, total 51 countries. However, Note: USA nor Russia were the initial signatories. France was not one of the 5 'powers' of the UN.
PCIJ (Permanent Court of International Justice)	01.06.1920	13.12.1920	30.1.1922	2	Replaced in 1946 by International Court of Justice
NATO - North Atlantic Treaty Organization	17.3.1948	01.06.1948	4.4.1949	1	Only the "Western countries, including Turkey are the members
ASEAN - Association of Southeast Asian Nations	01.01.1967		8.8.1967	0.5	Now 10 countries. China, Japan, South Korea, Australia, India, and New Zealand have trade agreements with ASEAN countries. Together they cover 45% of the world's population and about a third of the world's total GDP.

None of the organisations in the list have supranational powers that would be required if it were to combat existential risks successfully. The closest one is NATO with its article 5 that may engage a country against its own interest into a war, in which it would not have otherwise taken part. UN has two articles, 2 and 51, which guarantee the country's sovereignty. That is why it so ineffective, especially with the unanimity required for any Security Council resolutions. That is also why the International Criminal Court (of which the USA is not a member), cannot do anything to bring people responsible for genocide or crime wars to justice without the agreement of the government of a given country or a voluntary submission of the suspected criminal.

But even if we consider how minimalistic the restrictions on sovereign states were in the global organisations listed in the table, it still took a long time for them to become fully operational, e.g. 18 years for the Universal Declaration of Human Rights. Therefore, creating a new supranational organisation, which would most likely be a federation on par with the legal system similar to the USA or Germany, replacing the United Nations, is very unlikely in such a short time (about 10 years). But when one additionally considers the curtailment on certain rights and freedoms that might be necessary in certain circumstances,

like limiting sovereignty and some personal freedoms, then the chance of ever creating such an organisation that would include all countries is rather very small.

Therefore, the only other possibility to create a de facto World Government would be to adapt an existing organisation by the re-assignment of its scope and prerogative. To make this objective achievable within 10 years, it is almost certainly that it would not be joined by countries such as China, Russia, Saudi Arabia or perhaps even the USA, neither by the countries with deeply different values and interests. So, this 'partial' 'World Government' would have to co-exist with the countries outside this organization, which is a kind of an existential risk on its own. However, I believe Humanity has no other option and has to take this path.

Selecting a candidate for a Supranational Organization

Now, the question is who could do it? What we are discussing here is how to guide Humanity and protect our civilisation in the next 20 years, while it passes through probably the most dangerous period in its history. By then Humanity could reach the point when it may already be coexisting with Superintelligence, which would hopefully act in unison with us, helping to sort out our problems.

We are already late in creating such a large supranational organisation. Perhaps one of the reasons is that many of us still hope for the UN to take up such a role, or that the UN itself will be quickly transformed into such an organisation. After all, this is the organisation that should deal with existential risks in the first place. Unfortunately, this is also the organisation that indirectly **increases** the Humanity's overall existential risk by being almost totally ineffective in solving grave problems. UN may be trying to apply some solutions but quite often when it is too late, or a grave human, economic and ecological danger has already happened.

In most cases, UN is just incapable of solving the problems at all (e.g. Syria or Libya). The best evidence in my view is the creation by the UN of the International Strategy for Disaster Reduction (UNISDR) the organization that was established in 1999. It has had hardly any influence on any of the top ten existential risks listed earlier, apart from climate change and that was really managed by the International Panel on Climate Change (IPCC). The reason behind this is simple. To act successfully, such an organisation would need the powers far superior, which the UN has now. Since we know how embarrassingly ineffective the UN is, no wonder that the world is practically left alone to be gradually sucked into the funnel of ever-growing existential risks. Initially this

will be hardly visible but further into the future the risks will tend to combine. At that time there may be no organisation or country powerful enough to stop the demise of our civilisation.

Transforming UN into the World Government is also a futile hope, because of the very way the UN makes decisions – the unanimity voting. This is the same reason why the EU has been less effective than it could have been. Fortunately, the Lisbon Treaty has finally created a possibility for the European Council to vote using a qualified majority on more subject domains than ever before.

Since the existential risks can materialize at any time, e.g. pandemics due to laboratory-generated bugs being maliciously released into the open, we should have an organization that could act as *de facto* World Government right now, at the latest by about 2030. I have already indicated in the previous chapter that creating the World Government from scratch in such a short time, even if it only included most, rather than all, countries, is not feasible. Therefore, the only way that could happen is by transforming an existing organisation, or empower a single large country with supranational powers, into the World Government. That new global organisation would gradually substitute the UN, based on a fundamental review of human values discussed in the next part of this book.

So, we know now that it is unrealistic to create the World Government from scratch; it has to be formed from an existing organisation, whose scope and prerogatives must be greatly extended. But for such an organisation to be successful in mitigating existential risks it should have a mandate from **all of us** to act on our behalf. That of course will not happen. Even if there had been a world poll agreeing to establish it, that would certainly not come to fruition. Would those who hold power in autocratic or dictatorial regimes give up their privileges and introduce a democratic government as part of the World Government? Certainly not! It's a pity but that is our world and our civilisation and perhaps that's one of the reasons why we are in such an existential danger.

Therefore, the only plausible solution, is to assume from the very start that any supranational organisation that is to be created will only include the **majority** of nations but not all of them because there would simply be no agreement to do so. We see a similar scenario happening right now. The European Union wanted to go ahead and formalize a political Union, but the UK would never agree to it. That's why it has decided to exit from the EU. That has motivated some of the remaining 27 EU countries to move ahead towards a federal European Union, but those initial initiatives have been stalled by current events.

We must remind ourselves again that we need the World Government for one key reason – to mitigate the existential risks that Humanity faces. One of these existential risks is the Global Disorder. To mitigate this existential risk, we need to change entirely the way our civilization functions. That's why we need create the world-wide organization with significant powers. The migration wave of

2015 into the EU is the best example of how quickly national opinions change and become deeply entrenched. Therefore, we must look at a candidate organization that would be capable to reduce the existential risks in general, but in particular, its capability to minimize the global social and political disorder. Such an organization would have to improve dramatically the coherence of objectives between the member states, and from a global perspective.

Who has the best credential to become the de facto World Government? To make that assessment, we need to specify what would be the scope and prerogatives needed for such an organisation to be successful in mitigating existential risks, ignoring for now other objectives that it may have. I would suggest that such a check list should include the following questions:

1. Could the organisation execute supranational powers over a large part of the globe?
2. Does it have or will it soon have its own army and rescue services?
3. Will it be able to redefine human values that would become the foundation for the future new constitution and a legal system underpinning its operations?
4. Could it ensure very fast and co-ordinated response in emergency (potentially within hours)?
5. Does it have a large reserve of emergency supplies of food, seeds, etc.?
6. Does it have experience in dealing with large scale, global crises?
7. Does it have long-term experience in democracy and the rule of law, so that any decisions are made according to democratic rules?
8. Does it have enough resources, including financial, to deal with the current existential risks?
9. Is it very likely that it will be open to free and fair criticism and will it act on it?
10. Will it be able to adapt the way it works and introduce new laws very rapidly?
11. Does it have almost immediate access to best scientists and practitioners in every domain?
12. Does it have, or will it be able, to develop early warning system?
13. Could it create a very large civilization's refuge (a physical space in case of a catastrophic danger, i.e. huge caverns, or tunnels)?
14. Does it have and can it store large supplies of vaccines and medicines?
15. Is it, or will it be, capable of reducing nuclear proliferation?
16. Does it have or will it be capable of a strict oversight of molecular technologies?
17. Will it be able to fight populism with facts?

Now, that we know what capabilities such an organisation must have to mitigate existential risks, how would we then select the best candidate to play the role of the World Government. To select such an organisation, I have created a table

with 10 selection criteria for 10 organizations or large countries. I have tried to make the selection as objective as possible. 3 of the 10 criteria that I have used are completely objective: military power, territory size and GDP. The remaining 7 criteria are subjective, but that subjectivity is within a narrow margin, which over the 10 criteria could not make a big difference. The whole objective of this process is to select an organisation, which is likely to be one of the top three candidates, whatever the weights.

Weight	Justification for Selection criteria for the World governing organization	
10	Democratic institutions	This is the most important criteria because if we want to assure that we do not make things worse than they are now, then the nations that will surrender good part of their sovereignty must be assured that they will be governed within the best democratic system humanity has ever created
9	Respect for Human values	This is the second criteria in importance for two reasons. The organisation must be exemplary in its respect of human values and it has to carry out the process of redefining them for the upload to Superintelligence to make its risk as low for Humanity as possible
8	Military power	Any organization that will carry out such a role must be one of the most powerful in the world to withstand the threats from countries that will not be its member and carry out missions to minimize the risk to humans, such as Weaponized AI, or wars that could become global, or are of genocide type
7	Economic power	This is important because the organization must have enough resources to mitigate existential risks
6	Organizational capability	Essential when carrying out missions to eliminate threats from existential risks, such as nanotechnology
5	Response time to risk	The selected organization must be capable of very fast response to risk, sometimes within hours, i.e. nuclear war threat or artificial
4	Land mass	This is important to have available resources as well as creating spaces that may not be contaminated, e.g. biochemical risks
3	Experience in large programmes	Essential when carrying out missions to reduce existential risks, such as global socio-political risks
2	Versatility	The organisation which is to mitigate all kinds of risks endangering humanity must be very versatile and not for example have experience in the military field only
1	Neutrality, Objectivism	This is again important to assure cohesion of the organisation that will have powers to reduce freedom or sovereignty

The results are presented in the table below. As you can see, the organization that has come at the top is the **European Union**. It could be gradually transformed, initially embracing only a few countries, from the current confederation status into a full Federation, in a similar manner to the Eurozone expansion. It is already planning to take on new members, so in the next 10 years we may have other countries such as Ukraine or Georgia as members of this re-invented organization. I am expanding this subject in the subsequent chapters of the book.

Name of Organization or State	Risk Mitigation Capability Ranking (weighted)										
	Democ- ratic Insitu- tions	Respect for Human values	Military power	Econ- omic power	Organi- zational capab- ility	Resp- onse time to risk	Land mass	Experi- ence in large progr- ammes	Versa- tility	Neut- rality, Objec- tivism	Total Score (weight * capability)
Weight ----->	10	9	8	7	6	5	4	3	2	1	550
European Union	10	10	7	9	10	10	6	10	10	10	503
NATO	8	9	10	10	10	10	9	7	4	9	495
USA	9	9	9	8	9	9	7	9	9	9	480
Japan	10	10	3	6	9	9	1	5	4	9	391
Canada	10	10	4	4	9	9	4	3	2	10	388
Australia	10	10	3	2	9	9	3	1	3	10	358
United Nations	10	10	2	2	8	5	2	6	10	10	349
China	3	1	7	7	8	8	5	10	9	1	301
Russia	4	3	8	3	6	6	8	10	9	2	300
India	7	5	4	5	5	4	2	5	3	7	268

Let me now make a few comments on some countries and the assigned values. As you can see there is not a big difference between the first three countries but there is a big difference between the third (USA and the fourth (Japan). So, I will only make comments on the first three countries for each of the categories.

- **General remark.** If the scores are the same for an organization and a single country, then a country gets one point less because it is much more difficult to achieve a given rank, in an organization composed of many countries, than in a single country. Therefore, USA can get a maximum of 9 points.
- **Democracy:** NATO was scored lower because of Turkey (autocracy) and Albania, Bulgaria, Romania and Montenegro and Slovakia (all have too high corruption).
- **Human rights.** NATO scored lower because of Turkey (autocracy).
- **Military Power.** EU's military power score was the same as China's (because China is a single state). USA, the strongest power was scored 9 points because as a single state it gets 1 point less than a maximum 10).
- **Economic power.** No adjustment made.
- **Organizational capability.** USA scored 9 points because as a single state it gets 1 point less than an organisation.
- **Response time to risk.** USA scored 9 points because as a single state it gets 1 point less than an organisation.
- **Land Mass.** No adjustment made.
- **Experience in large programmes.** USA scored 9 points because as a single state it gets 1 point less than an organisation. NATO adjustment is due to experience in mainly military operations.
- **Versatility.** USA scored 9 points because as a single state it gets 1 point less than an organisation. NATO adjustment due to lack of versatility and focus on military operations only. That however may change in the future.

- **Neutrality and objectivism.** USA scored 9 points because as a single state it gets 1 point less than an organisation. NATO adjustment due to Turkey's operations in Syria and Iraq and autocracy of the regime.

In the end, even if NATO or the USA could have been chosen instead of the European Union, the whole process of changes that would have to be applied to any of these three organisations would be very similar. However, the changes to be applied to convert NATO or even more so, the USA, into a de facto World Government would have been much more difficult. In any case, I consider the EU as a kind of a strawman to see what kind of organizational and political changes the candidate for the potential World Government would have to go through.

Is the European Union our best hope for the World Government?

European Union came on top in the selection process for the organization that could have the best chance of mitigating existential risks. Let's remind ourselves that the key question this book has been trying to answer is which organisation is potentially the best one to control the risk stemming from AI, when it achieves the status of Superintelligence and ultimately becomes a Technological Singularity. But Superintelligence is only one of the existential risks that need to be mitigated. Therefore, any organisation that we choose to act on behalf of the whole Humanity must be capable of dealing with other risks too, including the Global Political and Social risks. The EU seems to fulfil these conditions best.

We will now look more closely at the EU's capabilities, its strengths and weaknesses and the scope of reforming EU so that it could start acting as the World Government. In general, the EU has quite a few features that are important for that task such as:

- Nearly uniform human values and legal system
- A wide spectrum of activities comparable with the UN
- A lot of experience in large international projects, like the accession of 10 eastern and central European countries on 1.5.2004.
- Significant financial and material resources
- A system extending beyond a typical confederation, with the president, the Government (the EU Commission), the prime-minister (the President of the EU) and the Parliament
- Dynamism. That may surprise some people and yet, there are very few other large organisations in the world that are as dynamic as the EU. Over the last

60 years the EU has been continuously adding new members, significantly changing the way it operates and continually distributing resources to poorer members on a very large scale.

- Ability to expand rapidly by integrating more countries, which are themselves significant global powers, such as Canada, Australia, and Japan, with which the EU has already signed wide-ranging treaties.

Anybody who may have doubts about the EU's capabilities to act as the future World Government should consider the impact of the recent EU General Data Protection Regulation (GDPR) regulation. It came in force on 25 May 2018. Its intention is to harmonize data privacy laws across the EU, to protect and empower all EU citizens' privacy of the data they provide to organizations and to reshape the way organizations across the region approach data privacy.

Formally, the law applies on the territory of the EU. However, in today's global economy, legal regulations applied by an organisation such as the EU impact countries, organisations, and citizens everywhere. The result of this regulation can be seen by anybody surfing the Internet. New GDPR procedures give Internet users a much better control on what happens with their private data. What is striking is not only the overwhelming adherence to the law by American companies and organizations but also the speed of introduction of these regulations.

This regulation does not impact existential risks, but it shows the EU's capability of swiftly and rigorously implementing laws and regulations that may in the future reduce existential risks.

When in 2010 the European Commission asked people about the EU citizenship identity, 62 per cent of people said they already felt like EU citizens (91). In August 2017, 68 per cent of the population (33,000 people surveyed across the EU) felt "they are a citizen of the EU". 56% of people across the continent were optimistic about the future of the EU in general – a rise in six points on the previous survey published in the autumn of 2016 (92).

So, why are people in the EU so optimistic barely a year after a bad experience with massive migration? Partially the sharp rise in optimism could be linked to the elections in the Netherlands and presidential elections in France, where Emmanuel Macron saw off a far-right challenger Marie Le Pen. The other factor could be the end of austerity in Portugal and the kick-start of the growth investment programme in the Southern European countries. But there are other less direct reasons to justify such cautious optimism and they lie rather in the political than economic domain, like putting some structural reforms to

managing the EU borders (Frontex agency) and moves towards closer integration sparked off by Brexit.

If life rejuvenation is successful, then most of the readers of this book will see the first day of the 22nd century. But to arrive there, we need to go through a stage of transition. We need to begin the process of federalization of the Planet. The EU has certainly some long-term problems it has to solve. We will cover them in greater detail further on when we will identify key areas that need to be transformed.

Chapter 2

Other candidates for the World Government

NATO

NATO would be the closest organisation that could take up the role of the World Government after the EU. It can act relatively fast, although not fast enough for some existential risks, especially in socio-political dimension. It is a great multinational organization under one command. However, it does not have the necessary experience in wider world problems, economic, social and legislature areas, and that may be of paramount importance.

Its scoring has been reduced in the domain of the respect for human rights especially in Turkey, and in the USA (death penalty and some problems with racism).

The United States

It is a great country and has some superb experience in organising large-scale programmes such as the Marshall Plan. It has shielded the West during cold war through NATO - the organisation that it was instrumental in creating. However, the main point is that its voters are primarily insularists, thinking mainly about America first, especially now under Trump. Equally important might be the problem with the need of significant revision of the US Constitution, especially articles regarding elections, gun control or death penalty (which reduced the USA's score in human rights category). So, it is doubtful the Congress, and even less so the Senate, would agree for America to play such a role. Therefore, the USA is unlikely to take on the role of de facto World Government in the foreseeable future. On the other hand, in 20 years a lot may happen in America, especially that we live in exponentially changing times.

United Nations

It has done a lot of good over the last 73 years. It is very complex and takes care of most of Humanity's needs. Its biggest contribution has been in providing immediate help to war-torn areas or regions affected by ecological or economic disaster, including famine. Another important area is health aid and guidance through the UN's WHO Agency, which could be a model for executing supranational powers even within the current system, e.g. Ebola crisis in West Africa in 2013-2016. Additionally, the UN has been very helpful in spreading education, basic hygiene as well as preserving the cultural sites of world heritage through the UNESCO Agency. Finally, especially in the last two decades, UN

has significantly improved the way it operates its economic assistance through the UNDP programme.

However, the biggest problem is that it acts extremely slowly and that for some of its decisions to be valid, unanimity voting must take place. That stems from a fundamental difference of some values between the Western world, China and Russia and it will not be overcome in the near future. Its executive powers, including military powers are very poor indeed. That is why it cannot solve some near existential risks, like getting rid of nuclear weapons, because of its current organizational shape e.g. the working of the Security Council.

China

China has some very positive strengths that could have been useful if it were to play the role of the World Government:

- It has the largest population on the planet – about 20%
- It has vast land mass although smaller than Russia.
- It has some values based on long tradition and culture that might be useful for such an organisation, like long-term thinking
- The nation has been used to extreme sacrifices to achieve its ultimate long-term goal. The war with Japan is one example. The second one is the current 30-year unparalleled economic expansion, which has also caused suffering, e.g. forced relocation of millions of people
- The country has been largely isolated from the world for two millennia but now is rapidly becoming a global power (it was the top world economy in 15th and 16th century, accounting for at least a quarter of the world's output).
- It is a military superpower and has delivered large infrastructure projects at unimaginable speed and organisational skills. They appear to be great organizers of super large projects that would be needed in case of a global crisis.

But China has not known democracy at all. Its system of values differs quite a lot from the western values. Therefore, it could not be a good candidate to lead Humanity out of an existential danger. On the other hand, some underlying values, which underpin the current system may be worth considering. They are approaching, in my view, the value system built by Lee Kuan Yew, the founding father of modern Singapore. Additionally, Chinese can make a great sacrifice for the right reason. So, in the absence of a 'western' style supranational organisation if it could not be built on time, China might act on behalf of all Humanity unilaterally. There could be a price ticket attached to this, e.g. supremacy of China over the rest of population, but Humanity might thus be preserved.

China is right now a benevolent dictatorship but the model of the Chinese autocratic system has some similarities with the Consensual Presidential Democracy that I proposing further down in the book (please, please, do not run away thinking that I am suggesting the installation of an autocratic system for the EU). There are similarities and differences:

- China's President has the powers, which the President of this new organisation would need
- Until March 2018, China used to run in effect a system of multi-presidency. The current president was elected for 5 years. He could have been re-elected for another 5 year-term (maximum two terms are allowed) as his predecessor was. He would then 'appoint' or groom his successor (who would have to be approved by the Politburo and the 3000 delegates of the People's Congress). Once he would have stepped down and his successor would have taken over, he would join a 'team of retired generals and officials', becoming the 'Supervisor' of that team and keeping a close watch on the current president. In effect it was a system of triumvirate presidency – the successor apparent, current president and the retired president. Unfortunately, in March 2018, the law was changed, and the current president Xi Jinping became the president for life.
- Of course, the Presidency system that I am proposing differs profoundly in how decisions are made (in an entirely transparent manner) and that strong powers of the president are controlled by equally powerful system of accountability and in the worst case, a quick removal of the president
- China's party system is based on gathering supporting votes, starting at the base level. The more votes one gets, the higher a person can go up in hierarchy. This system, if you consider the number of votes needed in China, is perhaps no less democratic (but only in this aspect) than the one in the US, when the key agenda of any candidate for a president is to collect maximum financial support
- There is a kind of a weighted voting system. In that system the weight is the role people play in the Chinese society. Party members' influence is greater than that of ordinary citizens. The weight in this case is the party rank. However, it could be fairly easily transformed into a weighted voting rights democracy. Actually, to some extent that is also practiced in western democracy under a different disguise. For example, in the UK, an average Conservative party constituency membership is about 2,300, of which about 150-250 members show up on the day of an MP's selection. However, many candidates are selected by the HQ, which usually calls to the Chairman of the local constituency, gently suggesting the 'right' candidate to be selected for an election. Once the candidate is selected and if he belongs to two main parties, his odds of winning a seat are better than 1/3. That's how the system works in the UK and the USA – the bastions of western democracy!

- China, if it were to be democratised in some distant future, then the modifications to the system would be relatively straightforward. The most difficult area for China to transform into a real democracy would be the sacrosanct divisions between the executive, legislative and judicial powers, which are still in a fairly good shape in western democracies.

In summary, although China did not come up as one of the key candidates to act as The World Government, it could take up such a role in a critical situation where the fate of the whole Humanity might be at stake - see Scenario 3 in the final Part of the book.

Russia

This country would have been a good candidate for these reasons:

- It has some values (e.g. patriotism, sacrifice for a greater good) that might have been useful
- It has great cultural tradition that would have been of great value in enriching the Constitution of Humanity in the area of art and culture
- It is a military superpower that would be needed in the hour of need
- It has vast resources – 17% of the earth’s land mass – invaluable for Humanity especially if the physical large areas (‘safe havens’) would have to be created

However, it has failed dismally in absolutely key areas:

- It lacks democratic tradition
- The transition to a more democratic system has stalled. Even if it is successful, it will be too late.
- The separation of legal, executive, and judicial powers is quite often illusory.

From our civilization’s perspective it is a pity that this great nation with superb culture and some family values that would significantly strengthen the pool of values of Humanity cannot be considered as a candidate to act as the World Government. If it had only been for the Russian people, then most of their values would probably be as good for the transfer to Superintelligence’s ethical code, as the values currently practiced in the Western world, which in some areas reach a decadent level (that’s why we have to renew them). However, it is the Russian autocratic government and lack of long democratic traditions that stand in the way of selecting Russia in the foreseeable future as an acting World Government. Unfortunately, it is the Russian government itself that may contribute significantly to Global Disorder, as described in Scenario 2 in the final Part of this book.



4

PART 4 CREATING THE EUROPEAN FEDERATION

Please note: You can find more detailed information on the subjects discussed in this Part, in Volume 2 of TRANSHUMANS – “Democracy for a Human Federation”

Chapter 1

What values for the European federation?

About values

If we want to create a federation that will initially act as a de facto World Government, we need to start with redefining our core values. But there is also a broader need for redefining human values – existential risks. Among those risks, which are most effected by human values is the risk linked to Superintelligence, which may become our most dangerous adversary. Thankfully, this has already been appreciated by those that are directly involved in creating AI, and ultimately the Superintelligence.

On 5th January 2017 at the Beneficial AI Conference at Asilomar, California top AI scientists met and defined 23 Asilomar Principles that by now have been signed by thousands of AI experts. These three principles directly apply to values and should be observed by all those involved in AI research and construction:

- **Principle 2:** What set of values should AI be aligned with, and what legal and ethical status should it have?
- **Principle 10:** Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviours can be assured to align with human values throughout their operation
- **Principle 11:** Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.

Values can also be defined as broad preferences concerning actions or outcomes. As such, values reflect a person's sense of right and wrong or what "ought" to be. "Equal rights for all", "Excellence deserves admiration", or "People should be treated with respect and dignity" are representations of values. There are several types of values such as ethical, ideological (religious, political), values, and aesthetic values. Values influence people's attitudes and behaviour.

Values also describe people's basic needs, such as freedom, dignity, or comfortable life. They can be seen as a hierarchy. At its bottom are **personal** values that an individual hold and which are usually a selection of values of one or more cultures. At the next level are **cultural** values shared by individuals of a given group or a territory. Finally, there are **universal** values shared by most people world-wide. I would also refer you to the Maslow's hierarchy of values that I discussed in previous chapters, which has the same foundation as this set

of values but at the second level instead of cultural, it has a level of safety and belonging (e.g. to a group) and thus to culture in general.

One would assume there is no difference between Universal Human Values and Universal Values of Humanity. But in my view, there is such a difference. It deals with two aspects of values – their scope and their change in time.

Under ‘scope’ people understand that all values relate to humans. However, nature, including all animals, is a collection of passive objects that cannot argue for their values to be respected. Therefore, we should extend the scope of values, including not just the values that pertain to humans, but also those ones that relates to nature. Therefore, they should be referred to as the values of Humanity, rather than human values.

Restricting universal values only to humans would not only exclude all animals but more importantly, the new species that may be born as a result of AI developments, culminating in Superintelligence. An exponential growth of capabilities and intelligence may quite quickly lead to the creation of new thinking and intelligent beings, including Superintelligence, potentially with its own consciousness. This is where the second element – the change in time of the applicability of a given value comes into account. Leaving the Universal Human Values statement as it is, i.e. static, would be in my view incorrect. Therefore, from now on, I will use the term Universal Values of Humanity.

About rights

There is considerable disagreement about what is meant precisely by the term rights. It has been used by different thinkers for different purposes, with different and sometimes opposing definitions. Therefore, a precise definition of rights is difficult and can be controversial.

For our purpose it is most important is to see the difference between **ethical fundamental values** and **ethical rights**. In simplest terms values give the context for their application in real life as rights to something that originates in a given ethical value. Looking from another perspective, rights are legal, social, or ethical principles of freedom to do something or an entitlement to something. They are the fundamental normative rules about what people are allowed to do, or what they have the right to expect from others in relationship with them, according to a legal or a social system.

Values are usually associated with cultures or groups within those cultures, as well as with belief systems, e.g. when we speak about religious values or family values. They usually form articles of the nation’s constitution. Rights on the other hand are most often linked to individuals and are most often converted in

a common law. This implies that rights can cross group-boundaries. A typical expression used nowadays, ‘human rights’ is a good illustration of this point. Human rights are thought to relate to individuals regardless of those individuals’ group affiliations.

Rights are regarded as ready-made mechanisms or products that can be immediately applied or embraced, often to counteract customs or values. Values, on the other hand, seem deeply ingrained in societies. Therefore, they require time and patience to be changed and if they are changed, the ramifications of such a change could be absolutely profound.

Key rights based on Universal Human Values

In modern history there are three important documents that define human rights:

- UN Declaration of Human Rights
- European Convention of Human Rights
- Charter of Fundamental Human Rights of the EU

These are probably the most ‘humanistic’ documents produced by our civilization. They go far beyond the French Revolution’s key triple values *liberté, égalité, fraternité*. All three documents do not differ significantly. However, today we have two equal legal binding texts and two corresponding European courts. The first one is the European Court of Human Rights (ECtHR), which makes judgments on cases in 52 EU and non-EU European countries, based on the European Convention of Human Rights. The second one is the European Court of Justice (ECJ), which provides judgments based on the Charter of Fundamental Human Rights for the EU countries only. The current set of human rights in that Charter includes 50 specific rights, each covered by a separate article and all derived from these 12 human values:

1. Freedom
2. Democracy
3. Equality
4. Justice & the rule of law
5. Human dignity
6. Social Solidarity
7. Tolerance
8. Life
9. Peace
10. National Security
11. Family Safety
12. Nature & Beauty

Balancing rights and responsibilities

There is not a single word in the UN Universal Declaration of Human Rights that would define our **responsibilities**? We need this, to balance citizens' rights. These include, for example, protection of the environment and adhering to ethical standards that would be beyond the legal system, such as good neighbourhood behaviour, not abusing the benefits system, aligning oneself with the country's values ("waving the flag") or participation in elections.

We would also need to redefine the coexistence of the majority and minorities in a society and their respective rights. What are the limits imposed by the majority on the minorities? After all, the rule of majority reflects only one aspect of justice that may create at the same time other injustices, as many other democratic principles do.

These are all the questions that need to be answered when redefining the Universal Values of Humanity. If the EU starts acting at some stage as the European Federation, then it must consider the questions of values from the point of view of the whole Humanity. These redefined Universal Values of Humanity should be the basis for the new EU Constitution, which will effectively become the Constitution of Humanity. The reason for that is that there is probably no other practical way to get the agreement on the Universal Values of Humanity by all nations, e.g. members of the UN. Just the very act of establishing a new body that would have created and approved such a Constitution would have taken many years. Even if such a new body to redefine Universal Human Values is somehow established, there could be no realistic prospect that it would ever come to an agreement because it is precisely the difference in the interpretation and more importantly, practice of some core human values, such as freedom, which would make such an agreement impossible to ratify by all nations. That, by the same token, also excludes the existing organisations such as the International Court of Justice.

Implementation of rights and their maintenance over time has a price tag attached both in monetary terms as well as in keeping the ethical balance. That almost always means that somebody's right is someone else's responsibility to ensure that such a right is observed. For example, my child has the right to be properly fed and clothed and it is my responsibility to make it happen. People have the right for emergency hospital care in case of an accident, and this is our joint responsibility to pay taxes due, to ensure that such a right can be materialized.

The overwhelming focus on human rights has created an unhealthy imbalance by ignoring human responsibilities. We see it quite often in courts across the EU countries, when an offender seems to have more rights than a victim. It is clear

evidence of how sensible liberal values have led to so called political correctness, seriously undermining the political and social stability. The pendulum of liberalism may have shifted too far towards the rights.

Therefore, similarly, as for the human rights, there should also be a complimentary list of Universal Responsibilities of Humanity. Let me just to give you one example for Freedom. We all take it for granted that we live in peace and in relative safety within the borders of Europe. But there must be someone who delivers peace. It is the army and police who do it. In most EU countries, compulsory military service was abolished years ago. The result of this can be seen in the way young people behave. It is great to see them enjoy such a wonderful peaceful life. But this is like giving a little child a toy. A child is unaware of what it may cost their parents. Young people are not even taught at school that freedom requires contribution both in money (taxes) and in kind (e.g. serving in the army).

Chapter 2

No Federation Without a Reformed Democracy

The principles of democracy

On the scale of 0 (the worst) to 1 (the best), the EU and most western democracies scored quite well in 2015 poll on observing democratic principles. On clean elections they scored about 0.85 as opposed to China (0) or Russia (0.42). Similarly, on fundamental rights the western democracies scored 0.82, and even more (0.92 for Scandinavian countries) versus China (0.42) and Russia (0.41) with Sudan at the bottom of the scale (0.29). So, it looks that at least in Europe democracy is OK. Is it?

People feel that they have democracy because they have free elections. But over 200 years ago Jean-Jacques Rousseau commented on the British elections, saying they were no guarantee of liberty: “The people of England deceive themselves when they fancy they are free; they are so, in fact, only during the election of members of parliament: for, as soon as a new one is elected, they are again in chains, and are nothing.”

We seem to forget that elections are just a means of taking part in democracy, whereas we see it as an absolute, fundamental doctrine that has an intrinsic value itself. True, elections can lead to change of government, resulting from public electoral debates. However, if you look closely you can see that the whole process is tightly controlled by teams of professional experts supporting the competing parties who feed the media and the electorate using most sophisticated socio-political techniques of persuasion. This is what is now called post-democracy.

Italy was probably one of the best illustrations of post-democracy under the presidency of Silvio Berlusconi, who went as far as creating laws that protected him. And all that happened in the democratic European Union, without any serious debate by the European Council and no financial or political consequences for Italy.

Such potential problems with democracy and rule by the majority were considered by a number of Enlightenment philosophers such as already mentioned Jean Jacques Rousseau or political theorists such as Alexis de Tocqueville. The latter one discusses the pros and cons of democracy in his “Problems of Democracy”. He argues that democracy is the ‘tyranny of the

majority’, repeating the claim first used by Aristotle against the Athenian democracy and also by other philosophers such as John Stuart Mill in his famous 1859 book “On Liberty”.

That does not mean that Tocqueville opposes democracy as such, but rather the excess of power held by any group or individual. One of his observations was that if a decision taken by a majority, which might have not properly understood the issue (think about Brexit), is evidently wrong and harmful to all, a minority that lost in the elections cannot turn to any public legislative body or executive power to correct that decision. People who may think democracy always provides sensible and just outcome should consider these facts. Hitler, Mussolini, and Peron were elected democratically but only once when they took over the power.

The tyranny of majority can be best illustrated by the following hypothetical situation. There are three people at a café. When it’s time to pay, two of them agree that the third one is to pay. Is it fair? This is not so far away from reality if we consider that in most developed countries about half of the electorate does not pay any taxes. The tendency is that more and more people do not pay anything into a common pool of resources to safeguard the interests of all. For example, in the UK in just 8 years (2008-2016) the number of people paying no tax has increased by over 10% (from 33.4% to 43.8%) What interest would these no-tax-paying voters have to change that situation? (93)

What it says is that the main problem with democracy is nature itself. But not the human nature. Are you surprised? At the rock bottom, we as humans are driven by Darwin’s maxim “survival of the fittest”, in which efficiency of acquiring the necessary resources for survival (i.e. using the lowest level of energy) plays a key role. In this case it means: ‘why should I pay for something with the money for which I have to work hard (and spend energy), why I can get it free. This is a typical, almost an animalistic, behaviour. Quite often it works well for an individual. But never for a group, such as a nation. We have ethics, a society, democracy, and the rule of law, which overrides this very early evolutionary trait with a new one (a meme). It is called social solidarity, or simply co-operation. This is the best way to thrive in evolutionary sense, because it creates the highest chance of survival of a group and propagating its genes into the future.

Democracy to work properly, especially a direct democracy, requires much better-informed voters, who would realize the profound impact that unbiased information sometimes has about a party they wish to support. The rapid turn from a liberal democracy in many Western countries to far right and near-autocratic governments is not accidental. We have plenty of recent examples. Turkey, which until about 2010 was seriously considered as a future full member

of the EU, is now an autocratic regime. France, Italy and even Germany were all under a threat of the political power being taken over by extreme right politicians using populist promises to win votes. In Hungary and Poland, we actually have far right governments. Russia is, and probably will remain, for the next generation an autocratic, if not a dictatorial, regime. And on top of that, democracy is at a dangerous point even in the oldest democratic countries such as the UK and the USA.

So, do we have a crisis of democracy? Well, just look around. Value crisis begins when the practice of members of a society starts deviating from the values, they hold dear. When there is a general acceptance of corrupt practices and unethical activities, then the society as a whole is in value crisis. It creates a new norm of tolerating dishonesty, lies and immoral behaviour. In the contemporary world, there has been a gradual deterioration of moral and ethical norms. It is this deterioration in upholding values that are the foundation of moral and social conduct that lie at the core of the current crisis of democracy.

Therefore, we should re-define what a democratic system is and how it should be practiced to make it consistent with the core Universal Values of Humanity discussed earlier. They should form key principles of a new type of democracy.

Consensual Presidential Democracy

How can we then cure the faults in democracy? I believe we need to ask some very difficult questions regarding democracy, which may undermine our current understanding and acceptance of democratic principles present in various constitutional and legal systems world-wide. The challenges we face today are in some way similar to those ones the world was facing when it was entering the Second World War. If you imagine what it was like in 1939 ,when Europe had to fight the most horrible war in history, or like in 1940 when Britain was preparing to defend itself with all available means, then it may be easier for you to understand the necessary measures that may have to be taken to save us all.

Therefore, I have proposed some improvements to a democratic system, most of which could be implemented straight away, if the politicians had the will power to do so. Since most of these reforms, and similar proposals, go against the politician's personal interests, they may probably only be implemented under duress (social revolution) or when one of the existential risks materialises. However, we should try to advance as many of these changes proposed here, or by other researchers, as soon as possible to reduce the scope of democratic reforms that may be needed when we will be forced to do so. This includes tasks related to saving Humanity from various existential risks.

I have called the proposed new system of democracy - **Consensual Presidential Democracy (CPD)**. This is a system of democracy aimed at governing with maximum consensus, where the voice of the 'losing' minority is always considered. *It is described in detail in Volume 2 of Posthumans – “Democracy for a Human Federation”.*

It gives the President exceptionally strong powers against the strongest accountability and recall procedures, to enable him to play a crucial role as a conciliator and moderator between two opposing parties, each represented by one Vice President. This system deepens the separation of legislative and executive powers by having a technocratic government. It also has the widest representation of the electorate, where the representatives to the Parliament are elected using a combined First Past the Post and the Two Rounds System of weighted voting, and where the second chamber of the parliament is elected using a Sortition system (a random selection of representatives).

The overall assumption underlying Consensual Presidential Democracy is that we can only survive the extremely dangerous transition to post-Superintelligence period, if we work more closely together, which means a gradual federalization of the whole world. We must be acting as one planetary civilization, as Humanity. That is why Consensual Presidential Democracy is built on the following principles:

1. The political representatives are elected in such a way that they represent most closely the will of the majority but do not at the same time ignore the needs of the minority. The 'losing' minority is not being trampled down but always heard and all decisions are taken by consensus rather than being enforced
2. Politicians are mainly elected to pass laws and oversee the execution of the law of the land but do not play any part in the execution of these laws
3. That leads to a much clearer separation of the legislative and executive powers when the government is run almost entirely by technocrats and specialists. Even if such a government has some weaknesses, it delivers overall better results than a government run by incompetent and quite often corrupt politicians
4. Deeper separation of legislative and executive powers will also insulate law makers from short-term political pressures and deliver better quality law
5. Any long-term decision, e.g. lasting a generation like in education, health, or infrastructure projects, would be decided by Citizens' Assemblies, to which representatives are selected randomly, and then approved by both Houses of the Parliament
6. The key role of this new style of democracy is first of all to assist Humanity in passing through the most dangerous period in its existence with minimum harm. It will also prepare Humanity to be governed like a Planetary Civilization, which at some stage will be ready for an eventual coexistence

- and a possible merger with Superintelligence, irrespective of the form it may take - mind uploading, transhumanism or any other way.
7. The role of the President is crucial in Consensual Presidential Democracy for two reasons:
 - a. The President has exceptionally strong powers in order to make, the most important decisions at the time of crisis very quickly, such as the emergence of an existential risk
 - b. The President plays a crucial role as a moderator and a conciliatory force when the majority and the minority parties cannot come to a consensus on their own
 8. Although there is a centralisation of key powers, the range of those powers is kept to an absolute minimum
 9. Therefore, most governing powers rest with the lowest level of decision making
 10. This consensual style of government, when applied to a Human Federation, also supports the redistribution of wealth on a global scale. This will be mainly financed by the future rapid growth in productivity, provided by an immense number of intelligent robots delivering to humans the world of affluence never experienced by humanity
 11. The governmental powers are far stronger than in any western democracy right now, to make the government more effective in fighting existential risks that can materialize within hours
 12. The voice of the governed is heard and acted on by bringing to account the governing through strong recall procedures applied throughout the whole term of the parliament and not just during the elections
 13. Strong powers of the president and the government are counterbalanced by an unprecedented level of scrutiny, transparency, and accountability

Such a new style of democracy will, in my view, have a better chance of supporting the new European Federation and indeed any other organization, or a state, that would take on the role of the World Government and withstand even more severe challenges than those to which the EU may be exposed in the near future.

Chapter 3

Federate or disintegrate

The missing word in the EU vocabulary

(You can find further details on the proposed EU federalisation in Volume 2 “Democracy for a Human Federation”).

There is one word missing in the EU vocabulary. It has been avoided in almost every political debate. That word is of course ‘federalism’. Instead, as is typical in the EU’s parlance, a euphemism was invented that has been used for the last 20 years. It is the ‘ever closer union’. ‘Federalism’ was the word, which Great Britain could never accept and that has finally led to Brexit.

From the EU’s perspective, Brexit was a mixed blessing. On one hand, the EU will finally have more freedom to move forward towards a closer integration since ‘the great nation constantly pressing on the brakes’ would no longer be there. On the other hand, it is a pity that Britain decided to fulfil its post-colonial dreams about playing a more important role on the world’s scene by being outside the EU. If rather than ‘sitting on the fence’, it would have instead got deeply involved in the EU reforms, then such reforms might have been implemented in a more balanced way, whilst not dampening the accelerated integration.

It is certain that the EU cannot stand still because it would have been a symptom of deep paralysis that might on its own lead to the EU’s disintegration. Here, I wholeheartedly agree with the EU Commission’s view that the EU badly needs more integration, not less. However, I would put a caveat: such integration needs to be much faster than at any time in the EU’s history, leaner and “smarter”, i.e. involving EU citizens much more directly, and controlling the process through well thought out phases.

It is certain that the EU cannot stand still because it would have been a symptom of deep paralysis that might on its own lead to the EU’s disintegration. Here, I wholeheartedly agree with the EU Commission’s view that the EU badly needs more integration, not less. However, I would put a caveat: such integration needs to be much faster than at any time in the EU’s history, leaner and “smarter”, i.e. involving EU citizens much more directly, and controlling the process through well thought out phases.

As you might have guessed, I have voted ‘remain’ in the EU referendum and am a strong supporter of a federal Europe, provided that such a federation will affect

our life as little as possible. Additionally, such a federation must simultaneously meet the criteria for fulfilling the role of the future World Government, since this would probably be our only real hope of mitigating Humanity's existential risks. I know this view goes absolutely against the grain of the current situation but I believe that in less than a decade such a federation must happen. Otherwise there will be no EU. How can this be achieved? Well, Brexit provides a welcome backdrop to future EU initiatives and makes any reforms more obvious than otherwise might have been the case.

In most general terms, such a federation would have to be a slim variant of the United States, mainly for cultural reasons. The social, cultural and language differences are far more profound in Europe and other countries that might join the Federation at some stage, such as Japan, than in the USA. Therefore, the 'federal layer' would have to be very shallow indeed and include the foreign policy, defence (as part of NATO) and economic policies, including common currency. The rest of the laws currently enacted in the EU should be devolved back to individual countries.

We need a federal Europe as quickly as possible. To achieve that, the current generation of politicians must be superseded by a new breed that will take higher risks, presenting to the electorate the situation as it is, and not selling populist ideas in order to be re-elected. On the other hand, despite a wave of pessimism in recent years about the EU's future, it looks as if the European Union has finally started to change the way it operates. It is encouraging to see some concrete steps that drive the European Union towards a much stronger integration, and possibly federation, such as the Bratislava Roadmap agreed in September 2016, the former President Juncker's State of the Union address in September 2017 or President Macron's vision of Europe he presented in the autumn 2017. But most recently, it may be the enhanced version of the Future of Europe Conference, initiated by President Macron and Chancellor Merkel in November 2019 – see further on.

Why to have a European Federation?

I hope I have answered that question in Part 3, chapter 1. We need the European Federation (EF) because it is the best candidate for acting as a de facto World Government, which would take control over global efforts to minimize existential risks. Apart from that we also need to have a credible and effective organisation capable of **halting the current fast and already dangerous path to global political, economic, and social disorder**. Finally, the EU itself can only survive if it moves towards a closer integration, which now means a full federalization.

The current legal status of the EU is ‘confederation’. Before we go any further, I think I would need to clarify what is the difference between a ‘confederation’ and a ‘federation’. The difference between a confederation and a federation is that the membership of the states in a confederation is voluntary, while the membership in a federation is not. Some nations, which started out as a confederation retained the word in their titles after officially becoming a federation, such as Switzerland. The United States of America was a confederation before it became a federation with the ratification of the current U.S. constitution in 1788. The set-up of the European Federation that I envisage is of the ‘federation’ type, like German Federal Republic.

There should be a new style of world governance in global politics led by the European Federation. The rapid technological changes and globalisation will bring about a new global order and new, existential risks such as fast spread of pandemics. This may require extending emergency powers when almost any country will have to give up at least temporarily some of its sovereignty, to an international organization, to overcome the existential risks, such as pandemics, that may require global action within hours. Current emergency powers are used very rarely and in very specific circumstances. These new “emergency” powers may have to be used more frequently because the combination of various global risks, which may convert into an existential risk. The very first thing the European Federation (EF) will need is its own Constitution, which could become a precursor of the Constitution of Humanity to emphasize the ultimate scope and aims of the EF covering not just the European population but progressively all people on the planet.

Options for converting EU into the European Federation

Let me restate again why we are discussing the subject of reforming the EU, which at some stage could act as de facto World Government. The most imminent, and one of the biggest risks for Humanity, is that our civilisation will just keep going on until the point, beyond which saving humans might be a futile effort. We have already concluded that a rational discussion among all world leaders, say within the UN, on forming such an organisation like the World Government would have been rather unthinkable and utterly unrealistic. That’s why I have suggested that the European Union (EU) has the greatest potential to be transformed first into a federation, which at some stage would start acting as a de facto World Government.

Therefore, the member states of the European Union must decide within the next few years if they want to become a federated state, which I have proposed calling the European Federation. Otherwise, the EU’s inherent inconsistencies and

inflexibilities originating from significant economic, social, and cultural differences (even within the same Christian culture) will gradually rapture its structure leading to its disintegration. That could start a period of political instability, which would almost inevitably be exploited by Russia, leading to European wars with most disastrous consequences.

The creation of the European Federation must be done in such a way that the new state will guarantee the former nations maximum flexibility and the widest possible scope for self-governance, i.e. that most decisions are taken at the lowest possible level of governance. This would minimize the tensions stemming from cultural and social differences. The only way to do it could be through the process, which I call “cantonization of Europe”. That is a reference to the way Switzerland has been governed for the last 200 years, giving the maximum level of self-governance to each of its 26 cantons. That would effectively lead to the creation of the European Federation composed of former EU states, but in which some of the larger regions (say over 5m population) would have an automatic right to statehood, either on their own, or by merging with neighbouring regions of other states.

The other objective for this new organisation should be to address the existential risks facing Humanity. It would thus have to stand up not only for the people of Europe, but at some stage also for the whole Humanity in the absence of a credible organization that would have ideally included all nations. The main benefit would be the increased safety of all of us, and in the long-term - a more humane and just civilization.

Existential risks force us to take extraordinary steps to save human species from extinction. Artificial Intelligence is one of the biggest risks. However, at the same time, it could still help us make a transition to the new époque of the coexistence of humans and Superintelligence, by re-designing and implementing the new world order. Since we have very limited time, we must rely on the best organizational, technological, and material solutions that we already have. The most feasible way forward is to entrust the fate of Humanity to a widely reformed and federated European Union.

Therefore, from that point of view alone the future European Federation must:

1. Become a credible and effective organisation that will be able to mitigate all man-made existential risks
2. Ensure global political, economic, and social order necessary to focus all efforts on our survival, since lack of a reasonable global stability may become, through combinatorial effects, an existential risk on its own

3. Fight inaction in the domain of preserving what is best in Humanity – its values, its rights, and intellectual and cultural assets, to avoid the emergence of a valueless Superintelligence, or equipped with dangerous values and objectives that in the end could destroy us all.

But then we must also look from a narrower, European perspective. Over the last 60 years the EU has achieved something unprecedented in the European history – peace, continuous economic growth, and relative social order. That pedigree and experience has probably tilted the odds for selecting the EU as the best candidate for expanding its role, and after its gradual transformation into the European Federation, serve wider aims, than just purely European objectives.

Irrespective of a particular route that the EU takes or is forced to take towards federalization (e.g. at the time of a super crisis), it should announce some immediate interim measures. This would give the EU some relative calm, necessary for implementing the future Federation. There are, in my view, at least three such commitments that should be made by the Council of Europe:

1. **The EU will start preparing a new Constitution** by establishing a Constitutional Convention. This could happen even before the end of 2020, if the Future of Europe conference changes its ultimate objective. It would give every member state some arguments for its domestic electorate that if things are not so good right now, they could be improved in the future. This can have a positive, calming effect, although undoubtedly, during the long processes of negotiating the new articles of the constitution, there could be some protests organized by populists' parties to retain the status quo or even derail the whole process altogether. I would think the process of agreeing the constitution should be used to explain to the EU electorate the new proposals discussed by the Convention, so that the proponents of the federated EU would have more time explaining all the implications.
2. **The strategic direction of the EU is an ever-closer union that may ultimately lead to the EU becoming a Federation.** However, the EU needs to be prepared for unexpected very serious events, like a super crisis in the Eurozone that might significantly accelerate that process. The current Covid-19 pandemic has made that argument even stronger. Therefore, a very severe post-pandemic economic, social, and political crisis may very quickly lead to establishing a separate zone for the Federated Eurozone to stop a potential existential danger for the EU. The remaining countries will all be expected to join the federated EU as soon as they are able to do so. However, in the meantime, they will remain in a separate, legally bound zone to the federated part of the EU
3. **No secession in any EU country will be approved or mediated by the EU country until a new Constitution is created.** This would immediately solve a lot of potentially explosive conflicts in the EU, such as Gibraltar, Northern

Ireland, or Padania. Over the next decade, it would also enable a peaceful and natural change of borders when nations are spread across two countries like the Catalans and Basques in the northern Spain and southern France, or Trentino-Alto Adige/South Tyrol in Italy/Austria.

However, as I mentioned above, the Covid-19 pandemic changes everything. Therefore, these initiatives may no longer be realistic. Instead, the EU may federate into the European Federation using two broad routes:

1. **Fast Track federalisation.** This would be a rudimentary federation, when only the most necessary functions would be federated, such as defence, security, foreign affairs, and the budget, while the remaining ones will be in a state of limbo or even chaos. Such a federalization may happen even within a year by using article 20 of the Lisbon Treaty, which gives the EU members (a minimum 9 members are required) 'an option to 'establish enhanced cooperation between themselves within the framework of the Union'. That is another euphemism, which means in plain language, that a group of a minimum 9 countries could establish a Federation, which as a whole would still be a member of the EU.

The Covid-19 pandemic has been a testing time for the whole EU and it will be even tougher in the post pandemic period. The €750 billion Recovery programme intended to keep the EU countries together may not be enough to meet the requirements of the Southern EU belt, such as Italy, Spain, Greece and also France. If the discontent in these, and some other countries, does not subside, and at least 9 such countries will exercise their right of invoking article 20, then because of the chaos, which may ensue, there may be no time for a long Constitutional Convention. Instead, such a federation would have a very basic initial legal set up while the Constitution might be ratified later on, say within two years.

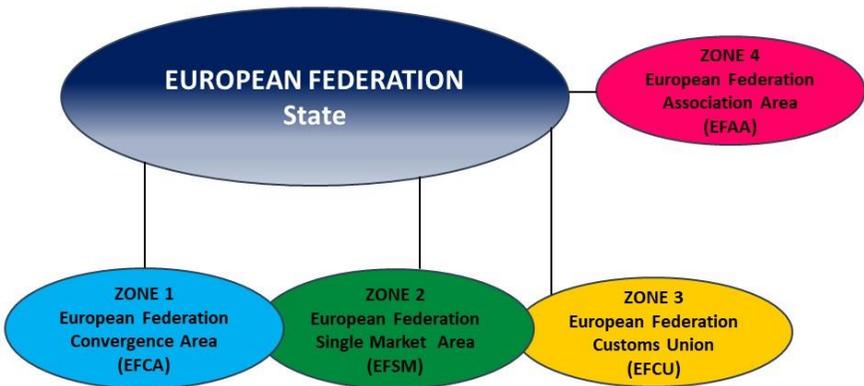
2. **Making a transition in a more orderly fashion.** As things stand right now, the EU is completely unprepared for a potential wave of shocks stemming from the rise of populism, tendencies towards autocracy in some countries, Euro crises, or referenda concerning regional independence. Therefore, an orderly transition to a federation is less likely than even a year ago. However, all is not lost yet. I mentioned earlier that on 27th November 2019 President Macron with a nod from Chancellor Merkel, initiated a 2-year programme called 'Future of Europe Conference'. Initially, it was to start on 1st January 2020, an unprecedented speed, and be completed by 2022. They even named the man to oversee the whole process – Guy Verhofstadt, a staunch federalist. But 27 European Council members, postponed the debate 'for later'. The date for the start of the Conference was postponed to the VE Day, 8th May 2020. However, because of the Covid-19 pandemic it was postponed again to autumn 2020.

The European Federation and its Subsidiary Zones

Assuming that the European Federation (EF) will be created, then on the day specified by its Constitution, all member states joining the EF will lose their sovereignty, and the EF will become a state on its own. On the same day, all national constitutions of the former Eurozone member states' will be replaced by new constitutions, in which their statehood will be changed to being a member of the EF. It will be a momentous event unprecedented in modern history. There is of course a possibility that some member states of the Eurozone or another group, if only part of the EU becomes federated, will not ratify the EF Constitution. These members will move to the EF Convergence Area (EFCA) zone (see below).

It is very likely that in any variant of federalization, most Eurozone countries will join the EF, although some may stay outside. They would then probably keep the Euro currency as a 'shadow currency', as it is now used in Kosovo and Montenegro. The Eurozone as such will simply cease to exist. The EF will become an ever-expanding core of the federal Europe, making the non-federated part of the EU progressively smaller. To enable such a gradual federalization, it seems that the best model for the EU to become a federation is to have the EF State with four subsidiary Zones.

The EUROPEAN FEDERATION



Although there are many other problems that would surface during the transition to EF, I would draw a conclusion that converting the Eurozone, or any other group of a minimum 9 EU members into the European Federation in the next few years. is definitely feasible. I would say, that in the current Covid-19 crisis it is almost inevitable. For many EU countries it will be a preferred option to reduce sovereign debt and many severe problems they may have in the post pandemic period

The future EF would also have to consider the impact of accepting new members on the cohesion of the Federation. Only the candidates, who would pass ‘the cohesion test’, should be allowed to start negotiations on accession to the EF. The cohesion criteria for acceptance would be the legal framework and actual practice of enforcing the Universal Values of Humanity in the candidate country. The second one would be the cultural similarity with the current members of the EU. The religion practiced overwhelmingly in the candidate country would be a particularly difficult and controversial subject. Therefore, some Muslim countries, such as Saudi Arabia or Iran (yes, we have to think 20-30 years ahead) would probably have little chance to become full members of the EF in the foreseeable future. However, the EF may enable special association membership with wider scope of privileges and obligations than at present for those members who would not pass the cohesion test. Turkey, Georgia, Lebanon, Ukraine are among about 30 countries that currently have such an agreement. Should the association criteria become more rigid, as they should be, some countries like Turkey, may not qualify to be associate members of the EF Customs Union and may be relegated to join Zone 4 – The European Federation Association Area (EFAA).

Zone 1 – European Federation Convergence Area (EFCA)

The setting up of Zone 1 may start in parallel to the transition of the Eurozone into the EF state. There will be no separate constitution for the EFCA and all the EFCA members would have to adopt in full the EF Constitution upon joining this zone.

Its initial joining members might be the EU member states that will be outside the Eurozone at the time of convergence of the EU into the EF. All members in this area must be the members of the single market and the customs union. It will be a transition area before joining the EF state. Any country in this area will have to join the EF state within a certain period, e.g. 5 years. Otherwise it may automatically be relegated to Zone 2.

Such a setup might initially create some tensions between the EF state and the EFCA members. Tensions might arise because of the perceived benefits of the membership of the EF state, not available to the EFCA member states and by the zone being considered a second-class membership. That is one of the reasons why President of the EU Council, Donald Tusk as well as the German chancellor Angela Merkel, have been opposed to the creation of the zones. But once the EF state has been created, that will only increase the motivation of the EFCA members to join it. The greatest benefit from this setup would be to show the EFCA members that the EF state works and any problems it might have, could be overcome. This is an optimistic assumption, I agree. It could also create some additional risks that are not present right now, and which might occur because of the changed political situation in the member countries of the EFCA. But at the same time, federalization of the Eurozone, or any other group of the EU countries into EF, and creating a separate EFCA zone for the remaining members could eliminate many more risks that threaten the unity of the EU today, like the looming Euro crisis. Therefore, the EU would have to take up such a risk.

Zone 2 - European Federation Single Market (EFSM)

The member states in this Zone will not have to accept the condition of an “ever closer union”, as it is required in the current Lisbon Treaty, i.e. they will not have to join the European Federation state. There could also be additional opt-outs for members of this zone, provided that they do not undermine the four freedoms of the EF state and the EFCA (free movement of goods, services, capital, and people). However, the EFSM members will have to be members of the Single Market and Customs Union. They will have the right to move to Zone 1, and eventually become full members of the EF, after meeting all the required criteria.

Zone 3 - European Federation Customs Union (EFCU)

When this Zone is formed, it will include all countries of the EU Customs Union, but which are not in the Single Market. Currently these are: Turkey, San Marino, Andorra, and some British territories. They will have to accept the jurisdiction of the European Court of Justice but not the EF Constitution.

The creation of the EFCU will streamline current Customs Union agreements and bring those countries closer to the EF, opening a possible path to joining the EF Single Market zone at some stage. This is why it is proposed that the members contribute 1% of their GDP to the EFCU budget, from which joint projects may be financed because it will create a stronger bond with the EF.

The setting up of this Zone may also start in parallel to the transition of the Eurozone into the EF. This zone will have its own Treaty, modelled on the current Customs Union Treaty and quite likely on the Comprehensive Economic and Trade Agreement (CETA) concluded between the EU and Canada in 2017.

Zone 4 - European Federation Association Area (EFAA)

We have to view the setting up of the EF as a major step towards creating a de facto World Government and the Human Federation. In the longer term, this Zone will probably be the most important one for converting the EF into a Human Federation, since this is a natural expansion path that ultimately could embrace all countries of the world. By creating the EFAA, the Federation would have the means to create much closer direct relationships with probably half of the countries of the world, shortly after the EF state has been founded, addressing the needs and aspirations of 2-3 billion people. This would make an additional natural pull for other countries to join the Federation, important if it wants to act as a de facto World Government. The reach of the Federation into other continents will at the same time help spread and re-enforce Universal Values of Humanity in many more countries.

The countries in this zone will have individual Association Agreements based on a Treaty with the EF, which will require the joining countries to make commitments to political, economic, trade, or human rights reforms in exchange for tariff-free access to some, or all EF markets (industrial goods, agricultural products, etc.), and financial or technical assistance. They will have to accept the jurisdiction of the European Court of Justice in matters related to their access to the EF market, but they will not have their Constitution aligned with the EF Constitution. Neither will they be the members of the Single Market nor Customs Union. In 2019, there were 24 countries that have signed the current Association Agreement, such as Israel, South Africa, Albania, the Ukraine, and 8 more countries, which are negotiating such an agreement with the EU. So, had the EF state been set up today, it would have had close relationships with nearly 70 countries.



5

PART 5
TOWARDS A HUMAN
FEDERATION

Chapter 1

An ever-expanding European Federation

The EU has a famous euphemistic expression when talking about the future federalisation: ‘an ever-closer union’. But the title of this section has to be taken literally. The Human Federation will not be set up from scratch. It will come into existence by a continuous expansion process, until it includes all countries of the world.

Let me remind you again, that the key reason why the world has to federate is to minimize existential risks. Two of these risks, developing by accident a malicious, Superintelligence, and the threat coming from climate change, both have a tipping point in about 2030. That is just a decade away. What it means is that after that year, controlling these threats may no longer be possible because of so called ‘runaway’ effect. In case of Superintelligence it may be the loss of control over its behaviour and a potential ability of setting its own goals. For climate change, the runaway effect relates to the feedback of a warmer air on the Siberian permafrost, Arctic, and Antarctic ice sheets, which will cause them to melt faster. This in turn will speed up the warming of the planet even faster, creating a loop of a continuous increase in the Earth’s average surface temperature.

The only way we can stop this ‘ticking bomb’ is by having a planetary organization that will act globally. Since we have just about a decade to have such a global organization in place, we must use all means available to avoid the worst to happen. That is why instead of creating a new organization from scratch, I have suggested to rely on an existing organization that would act as a de facto World Government. The EU appears to have the best chance to achieve that, after it has been federated.

The main condition for a successful continuous EF expansion and a later transition into the Human Federation is a well-functioning process of EF enlargement, since it needs to be completed within about one generation. The mechanism for such a continuous enlargement of the EF that I have proposed is based on the EF having four subsidiary Zones. In this respect, Zone 4 - European Federation Association Zone (EFAA) - is probably the most important one.

For a fast expansion of the EF, the accession criteria to Zone 4 must be relatively easy to fulfil. However, once a country becomes a member of that Zone and starts reaping significant benefits, there must be a stringent process of verifying that country’s adherence to the democratic principles and values set up by the EF. The whole process of a gradual transition of the European Federation into a

Human Federation can only function properly by implementing a new type of democracy, such as Consensual Presidential Democracy described earlier. This should be the basis for resetting the relationship between the nations, anchored to the Universal Values of Humanity, which best describe what it means to be a human being. This is the only way, in which the future HF may operate as a cohesive planetary civilization and how it can pass on its values to Superintelligence.

Therefore, the EFAA accession Treaty for Zone 4 members, must include certain principles, which will predetermine a candidate's suitability to join the European Federation, such as:

1. **Principles and values.** The joining members will have to strive to implement the Universal Values of Humanity, although they can get an exemption during the negotiation for some cultural or religious differences
2. **Respect for the rulings of the European Court of Justice.** Each member will have to accept the verdict of the ECJ in any aspects of its relationship with the EF
3. **Trading relationship with non-EF countries.** An EFAA member state can enter into trading relationship with a non-EF country without a prior agreement with the EF, unless it is also a member of the EF Customs Union
4. **Fiscal policy and common budget.** Members could voluntarily apply some fiscal policies of the EF
5. **Currency.** Each country uses its own currency. However, members may use Euro as a shadowing currency, taking of course all the risks that such a policy involves
6. **Opt-outs.** Members may declare opt-outs of any agreed policy of the EFAA unless it violates the principles of the EFAA Treaty. The ultimate arbiter will be the European Court of Justice.

By creating EFAA, the Federation will have the means to have much closer, direct relationships with more than half of the countries in the world by the end of this decade, addressing the needs and aspiration of hundreds of millions of people world-wide. Even today, there are over 70 countries, which have one of the 4 types of Agreements with the EU. This means they are either a full EU member, a member of the Customs Union, a Candidate member to the EU, or have signed an Association Agreement.

The EFAA Zone would make an additional natural pull for other countries wanting to join the Federation, important if the EF wants to act as a de facto World Government. The reach of the EF into other continents will at the same time help spread and re-enforce Universal Values of Humanity in many more countries. Gradually, after achieving a critical mass, and by applying the above principles, the EF will spread over the planet like an oil blob on the ocean, becoming one day, the Human Federation.

Chapter 2

Global Welfare State by 2040?

The world of abundance in just one generation

The unprecedented pace of exponential change may have either a positive or a negative impact on the long-term outcome for the human race. This largely depends on how we use the potential of such discoveries and innovation, like Artificial Intelligence.

As mentioned earlier, I assume that we shall have a fully mature Superintelligence by 2050. Based on that assumption, I have split the time between now and 2050 into three stages.

1. **Immature Superintelligence** - the first stage 2020–2030. It will start with the transformation of the EU into the EF in 2020'. I assume that by about 2030, the EF will already be an established federation, with all its institutions fully functional
2. **Twilight of Anthropocene** - the second stage 2030-2050. This is the period of rapid expansion of the EF until it transforms almost seamlessly into a Human Federation. It will then continue as for another decade as the only organization guiding humans to the period when we will be coexisting with Superintelligence.
3. **The Novacene Era** – Beyond 2050. This is the time of humans' coexistence with Superintelligence and a gradual evolution of humans into Posthumans.

I hope that by the start of stage 2, about 2030, we will have managed to eliminate most wars as a means of resolving conflicts and the world will behave more like a planetary civilization, co-operating in an increasingly greater harmony. By about 2040, when the EF may be formally converted into the HF, Humanity will be approaching its twilight. But what a twilight it may be!

So, far I have only stressed one reason for a rapid federalization of the world, to minimize existential threats, especially those, which may come from an unfriendly Superintelligence. That was a kind of stick, to change our behaviour, so that we can continue our evolutionary journey.

However, if we follow the world's transformation as I have been urging throughout this book, then we can also have a carrot. If we mature AI properly, so that with time it will become our benevolent partner, we will create not only a safer world, we will also achieve unimaginable prosperity for all people of the

world. This will be created by the mature Superintelligence, which will produce anything we want for a symbolic price, using billions of self-learning and self-creating robots. This may really be in reach within just one generation, i.e. by about 2050.

However, even before 2050, the expanding Human Federation will be able to create a Global Welfare State. Let me visualize what kind of life you may experience in not so distant future.

Education in Human Federation

In 2040, education at all levels has changed dramatically. In primary schools, traditional education has been almost completely replaced by the AI Assistants (one per classroom of 8). They perform the role of previous teachers but in addition they have in depth knowledge of every child's progress, since each child has an individual educational programme. The human teachers are still there but their main role is to teach children core human values and how they should be applied in life.

At secondary schools, almost the entire teaching programme is run by AI Assistants (1 per 4 students). Most secondary school pupils have brain implants, which have increased their brain capabilities exponentially. They communicate most of their progress wirelessly to their Assistants. However, similarly as in primary schools, they still have 'classic' lessons with human teachers at least once a week to teach them human values and enable them to meet their pals. They also have a sport day once a week. Those very few who don't have the brain implants, have an individual programme using a combined teacher-AI Assistant teaching method. Once a student passes an exam from any subject at least at 80% score level, he can then move to the next level. History and social subjects, like psychology, are in the main taught by human teachers, with AI Assistants checking the knowledge and assisting with any problems.

There is a minimum of two hours of history lessons weekly. Additionally, there are two hours a week of Human Federation Studies, of which 1 hour is dedicated to fighting fake media, populism, and xenophobia, and for discussing current events in the HF. There is a strong emphasis put on bringing up young people in the human values promoted by the HF.

All university studies are absolutely free. Students are assigned their own AI teachers, which teach a subject depending on the student's individual capabilities and aspirations. There are no formal exams, since the certificate on passing a subject depends on the entire work carried out by the students and dozens of ad hoc tests done under the supervision of AI examiners. Once a student has

achieved the required level of knowledge for a given subject, he is given the final degree from a university.

The future of work

The Technological Unemployment wave started a few years earlier than had been predicted because of the effects of the Covid-19 pandemic of 2020-21. The unemployment shot up to over 50% in some countries. The overall effect of Technological Unemployment was initially very severe indeed. Yes, over 160 new skills were created by 2030 as a well-known futurist, Thomas Frey, predicted in 2016. However, within two years of the first signs of the coming Technological Unemployment wave, it became obvious that there were far fewer new jobs, than the jobs lost and those jobs that were available, required rare skills. Millions of people became unemployed, resulting from continuous expansion of robotization and AI in general.

The lack of preparation in the old European Union countries for that entirely new type of unemployment was very obvious. That has sparked off serious social unrests in most countries. It was a very bumpy ride for employees and employers. Luckily, the unrest was quite quickly pacified by the introduction in almost all EU countries of either an unconditional Universal Basic Income or a Negative Income Tax. It has initially dented the budgets of some countries, but that was no longer a big issue after the federalization.

There is a compulsory job-sharing programme, which has almost immediately reduced unemployment. Every company that wants to make people redundant must immediately create two shared jobs or pay 50% tax on the salary paid for the position made redundant, for one year. In some cases, e.g. in companies, which operate as nearly fully robot-only production, there may be no possibility of job sharing. Such companies must pay an equivalent of 50% of one annual salary of the redundant human employee to the government's re-skilling fund.

The working week has been reduced to 15 hours, mainly as a consequence of the Technological Unemployment and there are plans to reduce it further to 12 hours in 2042. People normally work 3 days a week, which they can vary each month, by selecting the days at the beginning of each month for the next month.

Personal Finance

Growth of personal income has more than trebled in real terms in the last generation. That is having a remarkable impact on the changes of society's behavioural patterns. We are slowly moving up to the top of the Maslow's Pyramid of Needs; from the physiological and safety needs levels, to the levels

mainly pre-occupied with belonging, recreation, and inspiration to learn new subjects and practice unknown things (self-fulfilment needs). That happened almost naturally because of affluence and the availability of spare time. The only problem people have is with their personal safety (mainly cybercrime) and national security – Russia and China, which are now less hostile, and a bit more cooperative, but still not the countries people would trust.

There are no tax allowances, all personal income is taxable. The current rate of flat income tax is 15%.

Since 2030, there is a maximum value of assets that a person can have. It has been set this year at €100M and includes all assets, such as property, cash in bank, investment and shares, art objects and other personal belongings. Any excess of that amount is taxed at 100% but such a taxpayer may decide to allocate of up to 30% of taxed amount for charitable causes or social and scientific projects.

Every person, whether a child or an adult gets, an unconditional Universal Basic Income, which for an adult and a pensioner is an equivalent of 30% of an average personal income. This income counts towards the minimum living wage.

Every adult person may get a conditional Universal Supplementary Income, which for an adult and a pensioner is an equivalent of 30% of an average personal income. To get that income the recipient must fulfil certain conditions such as be in full time employment, or do a minimum number of voluntary work hours, or attend various education courses. This income counts towards the minimum living wage.

Every adult person must have by law, a guaranteed minimum income at the ‘poverty line’, which is 40% of an average personal income. This consists of unconditional Universal Basic Income (30%) and part of a conditional Universal Supplementary Income (also at 30%). That minimum income is in real terms equal to what over 20 years ago was the EU’s average personal annual income of about €30,000. However, to be eligible for such an income, a person must be in partial education or engaged in voluntary work, unless such a person is certified as incapacitated.

People, who are not complying with the condition to be engaged in partial education or engaged in voluntary work, do not receive a conditional Universal Supplementary Income (30%). If such a person has no house or flat, he is offered a free studio flat in the Government Funded Social Housing (GFSH), where he is given free meals and also any non-hospital medical care on site (including mental care). No people can be homeless by law and nobody can ‘sleep rough in the street’.

There is a minimum living wage that is an equivalent of 60% of the current average personal income and 20% above the minimum income. The Unconditional Universal Basic Income and conditional Universal Supplementary Income count towards a minimum wage. That means that any employer must pay net salary (after tax), which is worth at least 20% of the average personal income.

Young forever – rejuvenation medicine

Today, everyone takes it for granted that aging is reversible. The real breakthrough came in the early 2020'. Previously scientists were concentrating on altering a person's genome in such a way, that the body would not produce senescent cells ("zombie" cells that are half dead, best visible in the aged skin).

However, that appeared to be a cumbersome and a very risky treatment that might kill a treated person. Rather than genetically modifying a body, the scientists have been using a virus to encode genetic material, which "fine-tunes" the activity of certain genes, but leaves the genome alone. This leads to zombie cells being replaced by new cells. The effect is that many older people, well above 100, look now like their grandchildren. However, some of them are also Transhumans. Therefore, if a mature Superintelligence is finally delivered by the mid of this century, then it is highly likely, they will only live in their biological bodies for another 30-50 years, ultimately becoming Posthumans – an entirely digital mind.

Lifestyle

There has been a deep reflection on how to make human life as worthwhile as possible. After a period of about 10 years, in mid-2020', just as the most dangerous confrontation with Russia was subsiding, people begun very gradually returning to simpler forms of lifestyle. Initially, these were very small steps indeed, like limiting the use of plastic bottles or packaging and replacing them with more environmentally friendly solutions. Instead of using video phones, people started to see each other in person, especially when they finally realized how deeply their privacy has been compromised by digital media companies. That trend started just after the Covid-19 pandemic, which gave people time for reflection. Today, digital chatbots are passé. Meeting friends at cafés and at home is what counts. Tourism is booming, although most of these places can be seen and experienced using 5D holographic TV or special augmented reality equipment. People are becoming somewhat old-fashioned. It seems that the early digital experience was for many people like playing with new toys by children. Once they have played enough, they became bored.

Even family life seems to be regenerated, which probably stems from the same reasons as above. Since the average lifespan in the HF has now exceeded well over 100 years, in many families there are 4 or even 5 generations. Therefore, family reunions around birthday time can now be quite big events.

A high standard of life and plenty of free time has stimulated people's interests in the subjects, which they have never thought they would take up. Therefore, art, popular science courses, further education are the main element of their lives, since the working week is only 15 hours, soon to be cut down to 12 hours. Such interests and personal projects, if properly registered, such as genealogy research, painting lessons, or singing in choirs, can count towards the conditions necessary for receiving the Conditional Supplementary Basic Income.

Each person over the age of 13 can get their own Personal Artificial Intelligence Mentor. People wear it as a watch, which communicates with visual and audio receivers in a person's glasses, an implant in the eyes as lenses, an implant in the head or via any available wall display (although it is not recommended to be used outside home because of the lack of privacy). All information is stored remotely and is given top privacy level. It is given free of charge by the government, including the provision of associated services.

During the first few weeks, a Personal AI Mentor interviews the person in minute detail, makes a psychological profile and agrees with the person his long-term and short-term goals. It manages the person's all daily tasks and helps to complete some of them. The Mentor takes care of the person's all basic needs, including arranging any medical, mental, or other kind of assistance he may need with local authorities. The Mentor also arranges any work that a person is capable of performing, as well as any basic or even further education.

Initially people were very suspicious of such a powerful AI agent who knows more about them than they do themselves. However, today, most people do have them. They have become a very helpful way of enhancing people's life and making it far more interesting, enabling a lot of options and activities than otherwise would have not been possible.

Chapter 3

How to finance a Global Welfare State?

You may wonder how this Welfare State could be achieved by about 2040. What would be the source of such capital, which would support the needs of the EF itself, as well as the investment programmes envisaged for the members of the subsidiary zones? They may include more than half of the world's population by 2040. To answer this question, I present several sources of capital or significant cost reductions in most sectors that should be available from about 2030. These are:

1. **Much higher than predicted GDP growth.** This source is rather unusual, since it involves turning a problem (too low GDP growth) into an opportunity (much higher growth than would have been expected). This will be due to unprecedented growth of productivity driven by an exponential progress in technology, mainly in AI. Only very few economists share that view. Most are still entrenched in the old times, calculating growth as everything around us was happening at a linear, rather than exponential change.
2. **Tax rises.** This is the most typical source of finance for every government, although in this case even more important is the reason for doing that and its ultimate outcome (I explain it below)
3. **Significant fall in prices.** This could be a direct fall in prices (low inflation or even deflation) and indirect, through product substitution and product efficiency (a much greater value)
4. **Substantially lower cost of government** achieved by highest level of process automation and self-service
5. **Wealth cap – a significant redistribution of wealth** from extremely wealthy individuals by limiting the maximum value of assets a person can hold, e.g. over \$100M. Any excess would be 100% taxable, although an individual could direct up to 30% of the excess to a nominated charity
6. **AI-generated new type of wealth,** most of which would normally not be included in the GDP growth. This is the generator of wealth in every aspect. I immediately admit that a lot of the savings in this category will impact the fall in prices or will have already been in some way included in the previous sources. However, there would still be some 'leftovers' which are difficult to quantify and will emerge as new capabilities, never possible before, e.g. humanoid assistants providing elderly care in care homes.

Let me now put some detail on these proposals.

Higher than predicted GDP growth

All governments rely on budgets and budgetary forecast, especially for longer-term planning and making decisions on large infrastructure projects. And so, it would be with the EF. Before EF starts its ambitious socio-economic programmes, on the scale not ever seen before, it will have to look at its GDP growth projection, say 10 years ahead. If the expected GDP growth falls short of expectations, which has been mostly the case, it usually leads to significant budget deficits and the growth of national debt. However, what happens if the projection of GDP growth erred on the pessimistic side? That would mean that a number of potential projects that could have been included in the long-term national development plan, such as infrastructure projects, may have been totally excluded, as 'unaffordable'. Therefore, reasonably accurate GDP calculation is quite important. In the UK, for the last 10 years, the GDP projection is made by an independent body – Office of Budget Responsibility. In most countries, it is still being done by the government itself.

However, the inaccuracy of GDP calculation today may be a relatively small problem. In calculating GDP growth rate for 2040, we may be several times off the real figure. Let me support this supposition with some examples.

OECD in its long-term forecast assumes 3% annual growth rate for OECD countries between 2015 and 2040 (measured in Purchase Power Parity dollars, reflecting the real purchasing power of a basket of goods) (94). PWC assumes the GDP growth rate in developed economies over that period between 1.5 to 2.5% (95). Most of the long-term projected GDP growth ratio for developed countries oscillates around 2.5%. How credible is such a long-term growth rate? In my view it is not very credible. Who is right - orthodox GDP growth setters, or entrepreneurs and fringe economists? Right, in my view, are quite probably those people who do not have a vested interest in retaining the status quo. We face a similar situation today as regards the actuarial data that support the calculation of pension funds and their long-term liabilities. In most cases the data provided by actuaries is hardly credible. However, since the data is prepared and used by people who have a vested interest in pretending that everything is all right (that pension contributions are adequate to pay for future pensions), the contrarians have little chance to win the argument.

In many forecasts, almost everything depends on initial assumptions. One of the key assumptions is that change in all domains will broadly happen at the same pace as before. The reason for that is that there is simply no other data that could be the basis for assuming something entirely different as far as the GDP growth rate is concerned, i.e. rising suddenly much faster than linearly (as it must have because of the unprecedented AI-driven technological revolution).

Most economists still assume that the four components of productivity growth: labour, capital, technology (resources) and organization/entrepreneurship will grow largely unchanged as before. They seem not to appreciate how substantially the role of one of the growth factors – technology – has changed over the last decade and that its productivity growth is now reaching the level of Moore's law, i.e. doubling every 18 months. That's why the vast majority of economists still assume that the global GDP growth over the next 20-30 years will rise at about 2.5-3% p.a. That would mean that the world's GDP growth would barely double by 2040. Well, economists, governments and in general orthodox thinkers and planners have been spectacularly wrong on many occasions. Let me give you some examples:

- 2008 financial crisis. Nearly all major economists, or Harvard and MIT, have not predicted that catastrophic failure because they believed Milton Friedman's assertion that markets know best and they will re-adjust themselves
- 2015-2016 migration crisis in Europe. One of Germany's justifications for letting the migrants in was that Germany will need 10 million new employees by 2030. As I have shown earlier, the reverse is almost certain to happen, there may be more than 10 million Germans unemployed in 2030. In the context of Covid-19 post pandemic crisis and Technological Unemployment it may happen even earlier
- In Technology, Elon Musk with his Space-X Falcon 9 rocket, has within less than 10 years with his team achieved something that NASA or any other governmental organization were not able to do, i.e. to reduce the cost of payload vs. Saturn 5 rocket by about 20 times, among others, by re-using the same rocket (96).

That is why I am saying that the establishment is the least credible body to make correct judgments because of vested interests. The same goes for the GDP growth rate calculations. Take another example. In the most recent USA election, Mr Trump promised to repatriate largely manual jobs from China back to the USA, especially to the automotive industry. That was grossly misleading as it was simply impossible for many reasons. One of them is the fact that the American automotive industry has received since 2008 crisis, billions of dollars in direct or indirect aid, for restructuring their industry. The result is that it is now a significantly different industry with a much higher productivity than before. All thanks to the very latest technology. At Ford or GM, an hour of a robot costs now \$8 against £25 for a blue-collar worker. Even stiff custom duties proposed by President Trump could not change this situation. That kind of increased productivity like the one at Ford or GM, is still not properly being included in GDP model calculations, because it is like trying to hit the moving target, the data change too rapidly and is unstable.

But there are also other factors that point to GDP undervaluation. The global GDP is also undervalued because of purposeful action of some governments e.g. China, which undervalues its currency to boost export. Of course, if GDP is calculated in the same way, its growth rate will not be impacted. However, what will change is the real value (substance) that will be delivered, or how much more purchasing power a country will have. What I mean by this is that every year we consume more than would have been expected from the GDP growth alone. The “Economist” magazine has for many years calculated the GDP value using the number of hamburgers that can be purchased in any given country to reflect the meaning of real value (Purchasing Power Parity). In every country, the actual real value of goods delivered year by year is higher than the GDP growth would have indicated, and which partially forms the black economy (only a small part of it is by default included into GDP and in taxes).

Therefore, GDP growth will not be following the previous path. Instead, fuelled by relentless robotization and innovation, sometimes even exceeding exponential growth, (e.g. previously mentioned cost of artificial hamburger production fell 30,000 times in just three years) (97), GDP growth will be much faster even in developed economies.

This perception of more or less the same GDP growth is mainly due to missing the moment when change has passed the tipping point (called “knee of the curve” by economists) and from when change is accelerating exponentially. I believe we are just about that point, which means GDP growth will accelerate faster than orthodox economists envisage. For example, the whole agricultural sector in 20 years’ time will look entirely differently than today, because it will be cheaper to produce most food from basic chemicals, which you may think would be distasteful, but which actually is the other way around. Similar growth will be achieved in the productivity of various medicines (cutting down the time from a medicine discovery to the time it can be bought at a pharmacy), education, e.g. how quickly and how much more thoroughly, one would be able to educate an average person. We should, therefore, expect the GDP growth in real terms much more than just double by 2040. That will be an additional sizable income, which will allow financing of new social arrangements, like Universal Basic Income (UBI) and the Global Wealth Redistribution Fund.

Raising taxes to finance better life satisfaction

There is little correlation between higher taxes and higher level of happiness, or what I would prefer to call contentedness, as measured for example by the UN’s Human Development Index. Much more important is the government’s efficiency, the strength of democratic institutions, which is directly linked to the level of corruption. Taxes should be a means to an end and not the source for an

easier ride for the government to fulfil its sometimes entirely ideological commitments. And yet, the 2018 UN World Happiness Report ranks four Scandinavian countries at the top of the list, with Sweden (a very high taxation country) still making it into the first top ten (98).

Top 10 happiest countries, 2018 (2017 ranking in brackets)

1. Finland (5)
2. Norway (1)
3. Denmark (2)
4. Iceland (3)
5. Switzerland (4)
6. Netherlands (6)
7. Canada (7)
8. New Zealand (8)
9. Sweden (10)
10. Australia (9)

Why is Finland at the top of the list of the happiest people in the world (it was again at the top of that list in 2019), a country of 5.5 million people that only 150 years ago suffered Europe's last naturally caused famine? After all, GDP per capita in Finland is lower than even in its neighbouring Nordic countries and is much lower than that of the US. As all Nordic countries, they pay some of the highest taxes in the world (52%). But the Finns are good at converting wealth into wellbeing delivered by efficient and effective government. That's why just paying higher taxes does not necessarily correlate with life happiness. In Finland there is wide public support for higher taxes because people see them as investments in a good quality of life for all. The country has also been ranked as the most stable, the safest and best governed country in the world. It is among the least corrupt and the most socially progressive country with its police being the world's most trusted and its banks the soundest.

As you may have noticed, the Scandinavian system of government is for me one of the best overall in the world. Yes, Switzerland is an exception, as it is in many other aspects of government, having much lower taxation level and still being the 5th happiest country in the world. It is the country that did not have a war for 800 years, so its wealth has been accumulated for a very long time. Therefore, the case of Switzerland is not a good argument to claim that one can have a high standard of living, while also paying low taxes. If the government is efficient and effective, then higher taxes (at a certain level, not stifling the economy) would simply mean better economic and social personal outcome. That means the projects financed by the government, such as in transportation, are delivered on time and on budget.

In summary, the EF may initially need higher taxes to finance its Welfare State than the average EU tax level now. However, once it makes a transition to a Human Federation, the taxes at say 15% may be sufficient to pay for the world of abundance.

Significant fall in prices and a faster growth of real income

By about 2040, we will be in the period of continuously falling prices and a faster growth of real income, i.e. demonetization of the cost of living. This would mean that it will be cheaper and cheaper to meet people’s basic needs. All this will be driven by exponential growth in technological solutions and innovations in most sectors, leading to significant cost reduction in clothing, health care, housing, transportation, food, education, or entertainment.

Just think about this: the real value that is delivered to all of us, like Google applications, GPS, and other similar technology-originated services is not included in GDP because it is free! If you were to pay in 1982 for the facilities and services that you have on your mobile phone, then they would be worth, including inflation, about \$1million in 2020, not to mention a vastly superior quality, and unavailability of some services in 1982, like personal weather forecasting. Another example, video conferencing equipment in 1982 cost about \$250,000 (plus the actual cost of carrying out the video conferencing), which in 2020 money would be worth over \$600,000. Today, WhatsApp or Zoom applications that anybody can have on a smart phone is entirely free. Perhaps you only now realize that the phone you hold in your hand makes you a millionaire, as this table proves so clearly.

Dematerialization					
>\$900,000 worth of applications in a smart phone today					
Application	\$ (2011)	Original Device Name	Year*	MSRP	2011's \$
1. Video conferencing	free	Compression Labs VC	1982	\$250,000	\$586,904
2. GPS	free	TI NAVASTAR	1982	\$119,900	\$279,366
3. Digital voice recorder	free	SONY PCM	1978	\$2,500	\$8,687
4. Digital watch	free	Seiko 35SQ Astron	1969	\$1,250	\$7,716
5. 5 Mpixel camera	free	Canon RC-701	1986	\$3,000	\$6,201
6. Medical library	free	e.g. CONSULTANT	1987	Up to \$2,000	\$3,988
7. Video player	free	Toshiba V-8000	1981	\$1,245	\$3,103
8. Video camera	free	RCA CC010	1981	\$1,050	\$2,617
9. Music player	free	Sony CDP-101 CD player	1982	\$900	\$2,113
10. Encyclopedia	free	Compton's CD Encyclopedia	1989	\$750	\$1,370
11. Videogame console	free	Atari 2600	1977	\$199	\$744
Total	free				\$902,065

*Year of Launch

Source: Peter Diamandis and Steven Kotler: *Abundance: The Future Is Better Than You Think*, 2015

In 2040, it would be even more spectacular and likely that most of the things you appreciate will come free of charge. So, in 20 years' time we may be living in a world of abundance, at least twice as rich as today in real terms. Objects of desire previously beyond reach of an average consumer will become affordable. If we only assumed a very conservative doubling of GDP in real terms in 2040, that would make today's EU average personal net income of €20,000 per annum equivalent to what will then be the 'poverty line income'. That means, **everybody in 2040 would have as a minimum, today's EU average income in real terms, even if he would not work.**

Peter Diamandis in his article 'Why the Cost of Living is Poised to Plummet in the Next 20 Years' points out that in the U.S, in 2011, 33% of an average American's income was spent on housing, followed by 16% spent on transportation, 12% on food, 6% on healthcare, and 5% on entertainment. In other words, more than 75% of Americans' expenditure covers: Transportation, Food, Healthcare, Housing, Energy, Education and Entertainment (99). Let me use this data as a starting point for identifying the savings in 2040 in each of these sectors, beginning with food.

COST SAVINGS IN VARIOUS SECTORS in 2040

Food

The cost of food at home in the US has dropped by more than 50% in 40 years. In 2040 most food will be manufactured from basic organic compounds, including oil and some minerals as 'food' for specially selected stem cells for a given product such as beef, chicken, fish, or venison. That will be agriculture of the future, which will be very environmentally friendly, with no fertilizers, no pollution and occupying less than 1% of today's farmland and 1% of energy in developed countries. Additional gains will be made from significant reduction in cost of transport since most food will be produced locally. Even today, as mentioned earlier, 11 million hamburgers are produced from cultivated stem cells by just one manufacturer (still costs several times more than from a butcher's shop but is much healthier and tastier!). The whole process for a single stem cell to produce a hamburger takes about 9 weeks.

Healthcare

Healthcare can be split into four major categories:

- **Diagnostics.** AI has already shown that it is able to diagnose cancer patients better than the best doctors, image and diagnose pathology, look at genomics data and draw conclusions, and/or sort through gigabytes of phenotypic data. The real cost will be the cost of electricity and maintenance. The cost of

genome sequencing has been falling much faster than exponentially. In 10 years, it has fallen 100,000 times. In the UK by 2020 500,000 people will have their full genome sequenced and analysed (100). So, within the next 10 years all UK population will have their full genome scanned, enabling doctors routinely use a patient's genome data to individualize application of medicines, e.g. avoiding side effects, which are usually genome dependent.

- **Intervention and Surgery.** In the near future, the best surgeons in the world will be robots, and they'll be able to move with precision, and image a surgical field in high magnification. Each robotic surgeon will be able to call upon the data from millions of previous robotic surgeries, outperforming the most experienced human counterpart. Again, at negligible cost. By the way, it has been already happening for the last 17 years, with Da Vinci medical robot. It is only the cost that has prohibited its wider use. That cost fell to \$2M from \$2.5M in 17 years, very slowly and a typical robotic laparoscopy costs about \$3,500 more than carried out by a surgeon. However, the competition from the British made Versus medical robot, which is to enter operating theatres in 2018, will significantly lower an average cost of surgery, so that it will be at the same level as carried out by a human surgeon.
- **Chronic care and elderly care.** Taking care of the aged and chronically ill patients will again be carried out most efficiently by robots. Humans will still be needed to deliver psychological, mental, and palliative care.
- **Medicines and some human parts like kidneys or hearts.** They will be discovered and manufactured more efficiently by AIs, 3D printing machines, which will also do self-assembly.

Housing

- The building process is already being revolutionized by robots which build entire block of flats, and hence the scale will reduce the cost. The second reason is that people will only occasionally drive to work, working mostly from home and therefore will be able to live outside the city centres, which are far less expensive.

Transport (autonomous cars)

- These cars are becoming reality as I write this book. But the same will apply to public transport, which will be completely autonomous, and just more reliable (no strikes), safer, and significantly less expensive. Well before 2040, the commuting time will become the time to read, relax, sleep, watch a movie or have a meeting.

Energy

- Today, the cost of solar energy has dropped to about \$0.03/kWh, which is on par with the cost of electricity from some coal power stations. It will drop even further in the next few years by using almost twice more efficient solar panels.
- Scientists have been recently making significant breakthroughs in nuclear fusion, which is to deliver the first power station driven by its energy by about 2030.
- Most scientists also agree that by 2040-2050, the main energy source that we will use has not been invented yet (e.g. gravitons - think about the consequences of such an exponential change in energy production and use).

Education

- Education has already been demonetized in many respects, as most of the information you'd learn in school is available online for free. Here are some examples of free or nearly free Internet academies: Coursera, Khan Academy, Harvard, MIT, and Stanford. Each of them has thousands of hours of high-quality instruction online, available to anyone on the planet with an Internet connection. But this is just the beginning. As I mentioned earlier in Scenario 5, very soon the best professors in the world will be AI enabled Personal Teachers, which will know the exact abilities, needs, desires and knowledge of a student and teach them exactly what they need, in the best way and at the perfect rate. This is available right now. Think about the vast reduction of direct cost of education and indirect cost, fewer teachers (with entirely new skills) and smaller Universities.

Entertainment

- Entertainment such as video and gaming will cost close to nothing. Most of it will of course be in the form that we cannot even imagine, such as 5D hologram TV with synchronous sensitising, i.e. dispersing various smells, or creating surfaces of different texture that one might touch, etc.

Substantially lower cost of government

This is a significant area of cost savings that will directly impact the EF budget. Potential savings here are on the scale we cannot even imagine because a lot of these savings will be generated by the inventions that are not there yet. So, let me give you only some examples of those savings:

- All taxes and benefits will be collected and distributed with almost no human intervention. This will be the continuation of the process started in earnest about 2000 in all EU countries

- Cost of running Health care and Medical care will fall dramatically as indicated above
- The same is true about education, which will make a very intensive use of robots and AI education agents that will co-operate with teachers. There will be far fewer teachers needed with AI assistants. Those most needed will be behavioural and psychology specialists playing a very important part in the overall education
- The cost of the army will also be significantly reduced because the HF army will be miniscule in comparison to what it costs all countries today. There simply be no enemies to fight.

Wealth cap

The gap in wealth distribution between the richest and the poorest both in developed countries and world-wide is rising every year even faster than before, as indicated at the beginning of this chapter. That process cannot continue for two reasons. First of all, it is dangerous for the democratic system itself and is ethically unacceptable. It is dangerous even today for an individual, like Jeff Bezos, the resident of Amazon, to own assets worth the whole budget of a small country. This is the consequence of course of the crisis of capitalism and the way in which wealth, or sometimes pseudo wealth (i.e. trading in most derivatives) is generated.

Therefore, HF may introduce the wealth cap put on the maximum value of assets, which an individual can hold, perhaps \$100M. When introduced, such a one-off wealth tax should go back to the HF state, which will re-distribute it progressively over a few years to reduce the imbalance in wealth distribution.

Chapter 4

Global Wealth Redistribution Fund (GWRF)

Between a fish and a fishing rod

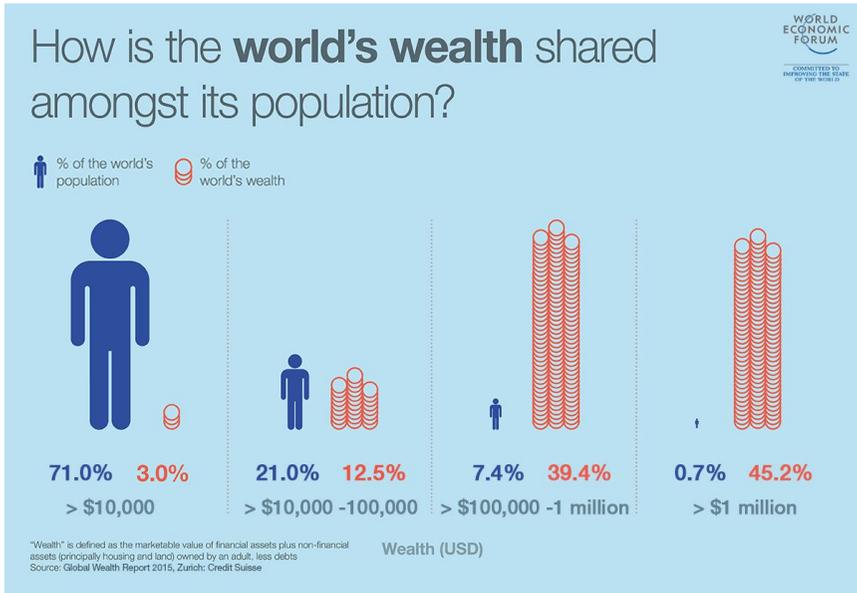
Let me now return to the question of the EF enlargement. The problem is that calling on other nations to join the EF and follow its democratic principles, or even trying to convince them that they must do it in order to save themselves from extinction, may not be enough. Until such threats become tangible and their negative impact is at least partially felt, it would not convince many countries to join the EF.

To achieve that, there must be something else. We must create, what I call, the Global Wealth Redistribution Fund (GWRF). The name may suggest that such a fund should exist to simply increase the GDP per head in the poorer countries, giving them a proverbial fish. But this is just the opposite, what I have in mind. What I am proposing is to give those countries, prospective members of the EF, a metaphorical gigantic fishing rod. What does that mean?

The only way, we can sensibly redistribute wealth is to have two strings attached to the help given to those countries. The first one is to ensure that almost all capital distributed to such countries must be invested in the required production facilities, infrastructure, education, and health sectors, so that they can themselves increase their income and indirectly their overall wellness. Any residual funds might be used as a stimulus for individual people to set up small workshops or enhance their skills. The other string that must be attached, as mentioned earlier, is the adherence to common set of values and democratic principles to achieve a more cohesive planetary civilization.

The scale and the urgency of the problem of wealth distribution is truly enormous. It proves that a global economic sustainability and world peace will be impossible without a significant change in the way we live together as citizens of the same planet. We really need a massive transfer of wealth largely from a Northern to a Southern hemisphere. I am fully aware of the complexities and difficulties of delivering such a momentous programme in the world, which today could not even agree on paying for CO2 reduction to lower the risk of a climatic catastrophe. The odds are heavily against such an optimistic view as I will present here. On the other hand, should we be incapable of resolving most of these issues by around 2030 then the world may face a bigger crisis than for example the WWII. We cannot create islands of sustainability. We cannot enjoy a sustainable life in an unsustainable world.

So, what is the scale of the problem? According to Global Wealth Report 2015, **less than 1% of the world's population owns nearly half of the global wealth** (financial and non-financial assets), whereas 75% of the world's population owns only 3% of global wealth.



What are we then doing to resolve such a wide gap in the world's wealth distribution? Well, in my view, we have not even started in earnest. The problem is not lack of initiatives. The problem is lack of money flowing from potential donors. The second problem is an effective allocation of funds. On 25 September 2015, the United Nations passed the resolution on Post 2015 Development Agenda, officially known as "Transforming our world: the 2030 Agenda for Sustainable Development". It is a broad intergovernmental agreement that acts as the successor to the Millennium Development Goals which involved 193 Member States. It contains 17 "Global Goals" with 169 targets.

Tony Czarnecki *Federate to Survive!*

UN 17 Sustainable Development Goals	
Goal 1	End poverty in all its forms everywhere
Goal 2	End hunger, achieve food security and improved nutrition and promote sustainable agriculture
Goal 3	Ensure healthy lives and promote well-being for all at all ages
Goal 4	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
Goal 5	Achieve gender equality and empower all women and girls
Goal 6	Ensure availability and sustainable management of water and sanitation for all
Goal 7	Ensure access to affordable, reliable, sustainable and modern energy for all
Goal 8	Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
Goal 9	Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
Goal 10	Reduce inequality within and among countries
Goal 11	Make cities and human settlements inclusive, safe, resilient and sustainable
Goal 12	Ensure sustainable consumption and production patterns
Goal 13	Take urgent action to combat climate change and its impacts*
Goal 14	Conserve and sustainably use the oceans, seas and marine resources for sustainable development
Goal 15	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
Goal 16	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
Goal 17	Strengthen the means of implementation and revitalize the global partnership for sustainable development

I believe that the SDG provides an excellent opportunity for the world to use this framework for much wider objectives, which would actually subsume the SDG. These are:

- **Create the wealth redistribution programme**, so that donor countries (mainly the Northern hemisphere) transfer over decades some of their wealth to those countries that need it most
- **Control mass economic migration in such a way that there will be no need to migrate**. That may require not only solving the poverty problem (mainly economic) but also environmental (scarcity of water) and political (civil and ethnic wars).

However, to achieve that, we need much more than the UN’s SDG agenda. We need a systemic global shift of wealth from richer to poorer countries. In theory, such a fund could be part of the UN Development Program (UNDP) but seeing the UN’s malaise in this area, I doubt it would attract funds on the scale that is needed. It clearly contrasts with the outstanding success of private funds such as Bill and Melinda Gates Foundation, whose vision is: “by giving people the tools to lead healthy, productive lives, we can help them lift themselves out of poverty”.

This is the key difference between how the UNDP and such a foundation works. For most of its existence, the UN funds were giving the poorer countries the proverbial fish, whereas private foundations gave them a fishing rod. Since 2000 that situation has improved at the UNDP, but the other crucial differences remain. These are efficiency, effectiveness of the projects and less corruptive

distribution of funds. It looks highly unlikely that the UNDP will change significantly, so that it could become the driver of such a wealth distribution. Therefore, it should be the EF, which would head such a programme, with additional injection of funds from other sources to finance the target GWRF projects, and if possible, co-ordinate wealth distribution.

I believe that such a large scale of wealth redistribution is the only realistic long-term solution for maintaining global peace and preparing Humanity for the new period of a planetary civilisation. Wealth distribution, if it is carried out on such scale, and if it follows the principles that I propose further on in this chapter, will achieve two objectives:

- It will virtually stop economic and climate originated global migration
- It will become a powerful and pragmatic mechanism for political change and instilling Universal Values of Humanity to all parts of the world.

I will expand these objectives in the sections below.

Stop economic and climate change-originated migration

Stopping massive migration by building a 7m high wall between Mexico and the United States that Donald Trump has initiated will not resolve the long-term problem. We need much more than just sending these economic migrants back home. They, and their countries, need help on site, both in building their own economy but also in education, health, infrastructure projects, and in building more just and equal societies there.

But some of today's richest countries, such as Britain, France, Spain, Portugal, Holland, or Belgium are also the former colonial powers. I know that from today's morality point of view, it would be very easy to judge harshly the people who inflicted such injustice and pain, like massive slavery and the robbery of resources. We can never excuse that although we can understand how these God-fearing people could at the same time treat their slaves as herds of animals.

Another example: how was it possible that the death penalty in Britain was only abolished in 1965? Morality changes and that is why we cannot judge our ancestors using today's morality, as the future generations should not do that to us. In any case, we cannot fully compensate the victims in these countries in any other way than providing material assistance. They deserve in principle to be helped to achieve by the end of this century, the same lifestyle as we have. The only way we can achieve that is through a programme such as the Global Wealth Redistribution Fund.

One can imagine two types of migration. The first type is the migration, which originates from natural catastrophes or man-made disasters, such as wars. The second one is an economic migration.

Let's consider the first type of migration. In the future, we may have natural disasters on epic scale – i.e. volcanoes or earthquakes, or several years of draught in Africa that would make large swathes of migrants fleeing to safe havens e.g. in Europe or the USA, from countries such as Africa, South America or South East Asia, which might become utterly uninhabitable. This will force tens or potentially hundreds of millions of people trying to escape their own homeland into the countries that have not been affected by natural disasters or simply less affected, and which would still have some resources untouched. It would be very difficult to propose a different solution for these people other than simply share with them whatever we have until the situation stabilizes and enables them a safe return home.

The second kind of migration is economic. Here we can help a lot. We have two broad options. The first one is to let them in with their families and make workplaces and homes in the host countries. We tried to do that in 2015. Initially, countries like Greece were flooded with migrants. They were supposed, under the EU Treaty, to offer them housing and any support they needed in Greece, since it was the first country that could offer migrants a safe haven. But very soon Greece ran out of resources. It was overwhelmed by the number and determination of the people, who had nothing to lose, perhaps escaping utter famine or civil wars in Syria or Iraq. Then Greece opened the borders to other EU member states, creating a chain reaction, as we all remember what happened in 2015 within the European Union. Over 1 million people entered the EU in just a few months. However, up to 60% of those migrants, as reported by “Independent” in January 2016, were considered economic migrants who took the opportunity and entered the EU with the wave of asylum seekers.

The problem created by that massive, mainly economic migration in a very short space of time, primarily to Germany, is with us till today. That was the key justification for various reforms postulated by populist movements, and which led to Brexit in Britain, right of centre governments in Austria, Poland, Hungary, and to effectively lost general election in Germany by Angela Merkel.

Now let's try to extrapolate that small wave of economic migration from the southern to northern hemisphere on a much larger scale involving tens of millions of people. How could it trigger wars and what could cause local wars to become global? As the EU example showed, opening borders to the neighbouring countries would create a chain of events. Very quickly those events would trigger off local wars, which through combinatorial factors that we were talking about before, could trigger off existential risks, such as natural

pandemics (lack of basic hygiene, spread of viruses such as Ebola etc.). Probably the only reason why the EU states 'invaded' by migrants in 2015 did not go to war with their neighbours was that they were members of the same European Union. It could have been an entirely different outcome, had the migrants tried to cross, for example, the Russian or the Ukrainian border.

People say, 'Democracy only works where there is a broad consensus about the distribution of wealth and power.' According to a number of surveys, only a proportion of the migrants flowing into Europe have fled from the immediate consequences of violence. Most of them came from places where there was no war. They just wanted better lives.

There are dozens of regions on our planet that are potentially a source of such massive migration. First of all, Africa comes to mind, where at the moment (apart from Sudan and Chad) people are driven out of their countries towards Europe in search of a better life. Then there is the entire Middle East, which may enter a new phase of war after the collapse of ISIS, namely the proxy wars between Iran and Saudi Arabia in Yemen, or Lebanon, Syria, and Iraq, not to mention Israel. And how about the Kurds spread over Iraq, Syria, Turkey, and Iran, who have declared their independence on 27th September 2017? Will they allow their aspirations for their state to be quashed for much longer? It is obvious that the old state borders of most Middle East countries are almost non-existent, and therefore these regions will fragment into stateless areas ruled initially by a local warlord. So, expect continuous bad or even very bad news from that area.

The current attitude of major powers and regional organizations such as EU or ASEAN is only making this risk even more credible. Mass economic migration (not to mention war-related wave of asylum seekers) now poses the most serious threat to Europe's stability since the end of the Cold War, and probably since the end of the Second World War. What can we do about it?

And this is the second option. The EU seems to be finally doing something that may be the beginning of a new policy in this area. In November 2017, French president Emmanuel Macron announced that the European Union and African Union would launch "concrete military and policing action" to rescue African migrants enslaved in Libya and arrest human traffickers. That was the result of the EU-pledged 'Marshall Plan' for Africa of €44 billion at a summit in Paris. So far, Tens of thousands of people were flown out of Libya to their native countries. It is a step in the right direction but far inadequate to the scale of effort needed to stop massive economic migration at bay. That could only be done by creating Global Wealth Redistribution Fund.

Instilling Universal Values of Humanity world-wide

I could easily imagine people reading this title and making a connection with the Spanish Armada's mission to spread Christianity in Americas. I could also see the associations with President Kennedy's Peace Corp programme, and perhaps some others, such as the EU Cohesion Programme, in which countries will get the EU help if they stick to the EU values. Well, I can only confirm that I would indeed wholeheartedly support the Global Wealth Redistribution Fund to be used in an effective way to exert some pressure to eliminate corruption, set up the rule of law and democratic principles based on the Universal Values of Humanity. This is nothing of the kind 'I know better what you need' in this requirement. We must all live in the world of peace, simply to survive as a species. We must all stick to the same principle of co-operation and mutual respect and understanding of what human life is about.

This brings me back to the title of this book: "Federate to Survive!". We now need more than a global co-operation. After all, co-operation and social solidarity is one of the 12 Universal Values of Humanity. We need a Human Federation to minimize existential threats, which may wipe out all of us within decades. If we want to minimize those risks we must get together as a planetary civilization, we must co-operate. Those who think otherwise may sometime be lucky and by acting in their own interest only, achieve a better outcome, usually in the short-term only. However, as soon as you sum up individual choices, where competitiveness is the name of the game, altogether those people will be worse off, had they co-operated with each other. Evolution has proven it over millions of years that for a species' survival the best approach is co-operation rather than competitiveness. We are in such a situation right now.

What does it mean in practice? There are richer countries within the EU that want to stop a transition to the European Federation, because they feel they will have to share their wealth with poorer members. However, over the last 15 years the poorer EU countries have already proven that thanks to the EU cohesion programme, their growth has been much faster, enabling the richer countries to increase their exports there. At the same time the EU as a whole has become a much more significant political bloc in the world, increasing the safety of all members. Safety is in numbers.

Now, expanding this example world-wide, it is also about the same Universal Values of Humanity, which in this case means sharing the wealth, so that those who are poorer will be able to sustain themselves, become more educated, more responsible and more resilient to various illnesses. For the richer, it would mean that they may avoid various threats and dangers that may come, for example, from mass migration from scorched lands of Africa. People deprived of basic food and water, may have no other choice than trying to cross the borders into

richer countries, like into Europe, even if many of them may die on the way. Such far larger migration waves like that of 2015, may happen a few times within this decade, destabilizing the world, should they combine with other threats, like wars and pandemics. That may lead to triggering existential risks and the extinction of all of us.

However, sharing wealth by giving money to those that need it most does not mean that those who give should not be concerned with how their money is spent. Therefore, we need to ensure that all money transferred in various projects will have a maximum positive effect and reach those who really need it. That is also part of fairness and mutual respect covered extensively in the Universal Human Values. Only then will a comprehensive decades-long programme of significant financial assistance in the form of Global Wealth Redistribution Fund (GWRF) make a real lasting change in making the world more peaceful and materially equal.

Setting up Global Wealth Redistribution Fund

For the world to achieve long-term prosperity and peace, wealth distribution is the most obvious and urgent task. The fastest way to do it is to emulate the current European Union cohesion programme, which is in fact a wealth distribution system to which each country donates 1.07% of its GDP (after the post Covid-10 when the 7-year budget is finally approved it will be much more). You may be surprised that something like this exists. But that is exactly the grand idea behind the European Project, especially successful, at least in economic terms, for the 10 countries that joined the EU on 1 May 2004. All of these countries had much faster GDP growth since then, even in the recent financial crisis period. All of them also have a national debt level far below the “old EU” countries. Similar projects led earlier to significant economic boost in countries like Ireland, Spain, Portugal, and also in Greece.

But if the GWRF programme is to work, then its scale should exceed any help or fund distribution the world has ever seen. It should be set up from the beginning as a decades-long continuous effort. The most natural way for the EF would be to act via the membership of the EF Association Area (Zone 4). The programme itself would be the magnet for countries to join that zone. Looking at the current EU association agreements it is obvious that the EU is already aiming in this direction, as each such agreement includes the development of political, trade, social, cultural and security links. Currently, there are over 30 such agreements plus over 50 trade agreements.

Once the EF has been established, most of these agreements will be turned into the EF Association Area’s agreements. Assuming that the rate of new countries

joining the EF's Zone 4 would be at least as fast as now, the total number of countries in the Federation's four Zones could reach over 130 by 2040. I include here, in the assumed EF Single Market Zone, the USA, India and Japan, the last two, have already completed their negotiations with the EU on the Free Trade Agreement, similar to the one signed in 2016 with Canada. That means in 2040 the EF would represent about 70% of the world's GDP and 60% of the global population.

This ambitious programme would be administered and mainly funded by the European Federation. At the moment, UN recommends that each country contributes 0.7% of its GDP to developing countries to deliver its 169 targets of SDG goals, signed by 193 countries. This is a non-binding commitment to the UN. Apart from the UK, only 5 countries (Sweden, Norway, Luxembourg, Denmark, and the United Arab Emirates) meet this obligation fully, with Britain being the only G7 country, where this commitment became the UK law in 2015. All the EU countries have pledged to meet that target till 2020.

I would suggest setting up of a fund such as the GWRP as soon as possible, even before the creation of the European Federation, say by the middle of the current EU Parliament, say by 2022. That may be needed and used as an emergency fund to control any large-scale migration wave or other unexpected events. It could be considered a kind of an insurance policy, as well as an ethical commitment to compensate at least partially for incredible suffering the colonial powers have inflicted on many, mainly African and Asian countries. As it happens, all those colonial powers are members of the EU (apart from Britain which has just left the club).

The UN SDG goals should be a starting point for funding the GWRP. I propose that until the creation of the European Federation, all EU countries contribute only 0.3% of their GDP to the UN SDG. The remaining 0.4% should go to the GWRP Fund, as it is informally being done right now, by channelling part of the money into the projects administered directly by the donor countries. The EU's Associate Members, such as Turkey or Georgia, should be encouraged to do the same in the future and sign an agreement with GWRP, or a similar fund, once it is established.

One of the requirements of the EU cohesion programme is that each recipient state also pays the same contribution to the programme as the donor countries (1.07% now). Additionally, once the projects have been approved by the EU or in future by the EF, the recipient country will have to contribute from 10% to 40% of its own capital to a particular project. However, because of the specificity of some recipient countries, which may be among the poorest in the world, the GWRP should work on a different basis than the EU cohesion programme. Therefore, I propose GWRP programme to be set up as follows:

- The GWRF should start its operation within the lifetime of this EU Parliament (by 2024)
- All donor countries would become members of the GWRF
- Any recipient country that wants to benefit from the fund must first sign an EU Association membership and later on become a member of the European Federation Association Area (Zone 4)
- Additionally, any recipient country that wants to benefit from the fund must also become a member of GWRF
- All members of the GWRF must pay their initial contribution to the GWRF of 0.4% of their GDP in the year of accession to the programme
- All members commit to increase their annual contributions to GWRF by 0.15% each year, until it reaches 2% of their annual GDP in 2030
- The recipient countries would additionally make a flat 15% contribution to the cost of each project carried out in their country

Should such fund be set up by 2024, then in 2030, the contributions in that year would amount to about 2% of GWRF members' GDP. That would be more than twice the current UN SDG goals. The combined EU's and some Associate members' GDP will be worth about \$30 trillion in 2030. That would make their GWRF's contribution worth about \$600 billion a year. If we consider the contributions from just the African countries, then their predicted GDP in 2030 is estimated at about \$5-6 trillion, which will make their contribution to the GWRF in that year (if they all join) worth about \$100 billion, plus an estimated \$50 billion contribution to the projects. Therefore, the total new capital injected to GWRF in 2030 could amount to at least \$800 billion because countries from other continents may also want to join, which would increase the overall pool of money available for investment.

However, to stop massive migration in its track, the future EF will have to do much more than that. During the first two years of the EF's operation the annual contributions should be frozen because of huge organizational changes and other challenges that the EF may face. However, for the period 2032-2040 the EF's contribution should gradually double to reach 3.5% in 2040 by continuing the annual contribution increase of 0.15%. When considering the GDP's much faster growth than ever before in that decade, and an increased number of new EF members, the injection of the new capital from the EF alone could reach about \$2.2 trillion plus estimated \$500 billion from the recipient countries in 2040. Thus, the total annual contribution from all GWRF members would be worth about \$2.7 trillion in 2040, equivalent to about 25% of the entire GDP of Africa in that year, although of course the GWRF projects will be spread on all continents, not just in Africa.

All projects would be closely monitored, as would be the expenditure from the current EU social cohesion fund, drastically reducing corruption, ensuring the project complete on time and on budget. It would also boost the world economy, especially in infrastructure projects, transferring not only the capital needed but also a wide range of skills to developing countries. The overall objective should be to raise the GDP per capita in poorer countries much faster than the world's average. For comparison, in 2016 the world's average GDP per capita was about \$10,000. In Africa, the poorest continent, it was \$1,809, i.e. 5.5 times lower than the global average. The projected growth for 2030 for the world's average GDP per capita is about \$14,000, i.e. 40% growth (101). For the transfer of wealth from the rich countries to poor countries to be meaningful, the GDP growth per capita in the poor countries, should be at least twice that fast (due to external help) as the world's average. Should that happen, then in 2030 the poorest countries' GDP per capita would be about 3.5 times lower than the global average, and about 2.5 times lower in 2040.

Such transfer of wealth should also minimize the risk of mass economic migration, although the effect will really be felt only from around 2030. As I mentioned many times in this book, the period till about 2040 may be quite dangerous for the world in many domains, including socio-economic area. One of the most difficult problems that will become apparent will be the Technological Unemployment.

This is a type of unemployment that will permanently eliminate not just jobs, but job types, including neurosurgeons, drivers, accountants, lawyers. Altogether over 100 job types may be eliminated by the end of this decade with a similar number of new job types introduced. However, the number of new actual jobs will be far fewer than the jobs gone.

It was expected that the Technological Unemployment will start in earnest in the second half of 2020'. However, Covid-19 pandemic again changed a lot. Many people that have been staying at home for several months, while being paid by governments, will not be able to return to their workplace, because their job types will be gone. Employers, who have suffered enormous losses have in that short time, re-invented the way products and services are delivered. This mostly means a significant investment in new technologies, such as robots, drones, self-checkouts in supermarkets etc. Since demand for such new technologies has now reached a critical mass, manufacturers will rapidly increase the production of robots and other technological solutions, which will reduce the unit price of these robots. This in turn will create additional demand, leading to a very fast reduction in job types. That's what Technological Unemployment will be about.

However, Technological Unemployment will affect the poorer countries significantly less than the more affluent ones. The poorer countries could even

play a role of lessening the negative effect of Technological Unemployment, if the funds like GWRP become significantly large. Potentially unemployed people in the richer countries could find work in the developing countries both on large infrastructure projects, and in education and apprenticeship programmes helping to upskill millions of people.

The donor countries' governments will have a formidable task to convince their taxpayers to pay higher taxes to finance such an ambitious programme. To make it easier, the financial commitment from the governments must be more visible, like it was with climate change 15 years ago. At that time, in many countries the public was informed once or twice a year, using the traffic lights information system on the progress of environmental clean-up and green energy uptake. Similarly, it should be done for the investment in the GWRP fund both for the donor and the recipient countries.

Particularly important could be the information on the effectiveness of the investment in the recipient countries. Although the money invested in any given country would come from many sources, it would still be of interest to general public to know what is happening with the money and how effectively it is being used. With some countries lagging significantly behind in delivery and effectiveness of the projects, public pressure to improve the situation might be quite helpful. In any case, transparency on such a large-scale programme will be a key factor motivating taxpayers to continue paying the country's contribution.

During elections, political parties should very clearly explain, avoiding political correctness, why such significant amount of money is being paid to poorer countries. The arguments mentioned earlier such as controlling mass migration by creating decent and safer living conditions in the recipient countries, or just long overdue moral obligation for the former colonial powers to help those countries, should be clearly articulated in party manifestos and the government's annual budgetary statements. Overall, supporting such a transfer of wealth is probably the most effective way of combating quite a few risks linked to Global Social Disorder. It is simply the best insurance policy.

We need to federate the world to survive existential threats, which may wipe out all of us within decades. If we want to minimize that risk we must get together as a planetary civilization, we must co-operate. Those who think otherwise may sometime be lucky and by acting in their own interest only, achieve a better outcome. However, as soon as you sum up millions of such individual choices, where competitiveness is the name of the game, all those millions together will be worse off, had they co-operated with each other. Evolution has proven it over millions of years that for a species' survival the best approach is co-operation rather than competitiveness. We are in such a situation right now.

What does it mean in practice? Let me start with the current situation in the EU first. There are richer countries within the EU that want to stop a transition to the European Federation, because they feel they will have to share their wealth with poorer members. However, over the last 15 years the poorer EU countries have already proven that thanks to the EU cohesion programme, their growth has been much faster, enabling the richer countries to increase their exports there. At the same time the EU as a whole has become a much more significant political bloc in the world, increasing the safety of all members. Safety is in numbers.

Now, expanding this example world-wide, it is also about sharing the wealth, so those who are poorer will be able to sustain themselves, become more educated, more responsible, and more resilient to various illnesses. For the richer, it would mean that they may avoid various threats and dangers that may come, for example, from mass migration from scorched lands of Africa. People deprived of basic food and water, may have no other choice than trying to cross the borders into richer countries, like into Europe, even if many of them may die on the way.

Such far larger migration waves like that of 2015, may happen a few times within this decade, destabilizing the world, should they combine with other threats, like wars and pandemics. That may lead to triggering existential risks and the extinction of all of us.

A Mini Welfare State for Developing Countries

I have already presented a broad vision of a Global Welfare State, which would of course also include the developing countries. However, these countries may need an urgent direct help to enable them simply to survive before we arrive at the world of plenty. So, what could we do in the meantime?

Most developed economies already have welfare states, which cover many sectors. The most common is of course the health service, such as free of charge British National Health Service, free education and extensive social services helping people in need. Welfare state varies across the world in its funding and generosity, but the principle is always the same: people going through periods of ill health or unemployment receive direct support. That support is usually in-kind and additionally in cash, e.g. child support allowance or winter fuel allowance. The assumption is that to address social ills, the government has to help the people in need, and that costs money.

The picture is very different for international aid, both development aid to economically underdeveloped countries, and emergency assistance after disasters or wars. Donor organisations and countries rarely hand out cash,

generally doing so only in emergency situations. Most humanitarian aid comes from a huge and complex network of organisations delivering many different kinds of aid.

Depending on your view on international help, some people would see it as a very positive humanitarian assistance delivered by strongly motivated people, quite often volunteers, to deliver expertise and well-targeted interventions to people who need it. Other people on the opposite political spectrum may instead talk about high salaries for international staff, incoherent investment programmes and long and expensive supply chains for delivering goods that could be supplied locally.

Without going too much into detail, I would like to sketch how some percentage of the money delivered to the recipient countries could be best spent helping directly individual people. You can call it a wish list combined with some ideas on how to minimize the suffering of the people who are at the absolute rock bottom of the world's personal income table, like those with GDP per capita of less than \$1,500 per annum (in International dollars).

Poorest countries in the world in 2016 ⁽¹⁰²⁾

Rank	Country	2016 International Dollars
1	Central African Republic	656
2	Democratic Republic of the Congo	784
3	Burundi	818
4	Liberia	882
5	Niger	1113
6	Malawi	1139
7	Mozambique	1228
8	Guinea	1271
9	Eritrea	1321
10	Madagascar	1504

Below is my wish list of measures that could be applied for the poorest countries financially supported by a fraction of the GWRF. My assumption is that it would fully work from the late 2020', although if the GWRF fund is created earlier, some elements of the programme proposed here could be introduced on a pilot basis in some countries as soon as the EF is created. Each recipient country would have to be a member of GWRF and the EF Association Area (Zone 4).

The cap for this kind of help for qualifying countries would be an annual GDP per capita equal to about \$1,500 in 2016, as in the table, although by say, 2030, the countries and the amounts, including the cap, may change significantly. All numbers are of course examples only. The programme would have to be properly

costed before it is introduced on a full scale. Some countries might have already implemented some of these suggestions, such as a minimum living wage, in which case, the money due would be invested in the country's projects. Here is my wish list:

1. Introduce unconditional Universal Basic Income (UBI) for all adult citizens. That income would equal 20% of the average wage but no more than \$500 per annum (valued at 2030 level). It would be funded by GWRP - 70% and the recipient country - 30%. There was already a similar programme ran in Namibia from January 2008-December 2009 by the Basic Income Grant (BIG) pilot project. All residents below the age of 60 received a Basic Income Grant of N\$100 (about US \$9) per person per month (about US \$40 per average household, i.e. about \$500 p.a. per household), without any conditions being attached. It produced some interesting evidence on the impact of UBI in developing countries as the final Report from that pilot project concludes: "The village school reported higher attendance rates. Children were better fed and more attentive. Police statistics showed a 36.5% drop in crime since the introduction of the grants. Poverty rates declined from 86% to 68% (97% to 43% when controlled for migration). Unemployment dropped as well, from 60% to 45%, and there was a 29% increase in average earned income, excluding the basic income grant. These results indicate that basic income grants can not only alleviate poverty in purely economic terms, but may also take the poor out of the poverty cycle, helping them find work, start their own businesses, and attend school" (103).
2. Introduce Conditional Universal Supplementary Benefit for those in employment for at least 16 hours a week. It would amount to 20% of country's average wage. with a cap of \$500 p.a.
3. Introduce minimum living wage, which would be at 60% of the average wage. Unconditional and conditional Income would count towards the living wage
4. In countries with very high unemployment introduce job sharing, where for every job shared, GWRP would pay extra 30% of the original wage of a full-time job
5. Introduce state pension financed from contributions (employee 30%, employer 30%, the state 40%)

This mini Welfare State might play an important role in reducing the tensions in the poorest countries and show them that a new way of life is opening to them right now.

Chapter 4

The European Federation becomes a Human Federation

When selecting the best candidate for a de facto World leader, there were three other potential rivals to the EU, namely NATO, USA, and China. However, the analysis I have carried out, clearly pointed out that the EF is the most suited organisation for such a transformation into the future Human Federation. Therefore, I would not expect that at some stage EF will pass on the control to another organisation such as the UN. If my assumption is right, it will rather be the EF taking over more and more functions from the UN, than the other way around. Given its current status, UN can only do what is possible within the existing legal and political constraints. But the world needs an organisation such as the EF right now, fulfilling the tasks that UN will probably never be able to deliver.

So, what next for the European Federation? The EF may almost seamlessly become a Human Federation by about 2040 because there may be fewer than a third of countries left outside the EF. In the final 10 years leading to 2050, the Human Federation may go through the process of consolidation and strengthening of its institutions. The fundamental document that the future Human Federation may inherit from the EF will be its Constitution. As any constitution, it will be a living document and will be amended as the needs dictate. The same relates to the EF's institutions. Most of them will be gradually adapted to a larger world-wide function, rather than the European role only. However, they will be some new institutions, which will most likely be set up on the basis of the tasks that the EF will have already been carrying out, but perhaps at a smaller scale.

The biggest task in the final stage of that transformation process will be the consolidation of the HF. The procedure for accepting and moving new members through the zones of the EF until they merge with the EF state, will have been well tried-out. Therefore, after 2040, within say a further 5 years, all the remaining countries may be joining the Human Federation. If it seems incredible then you must not forget that the world will continue to change at nearly an exponential rate. Additionally, there may be other reasons for such an accelerated, slightly chaotic phase in the HF. The world may already be experiencing in positive and negative ways the impact of the maturing Superintelligence. By about 2045, the humans' supremacy for making any decisions on this planet may be slowly waning. The evolution of the EF into a Human Federation will be nearly complete.

The Human Federation will have its Parliament and the World Government. There are many options for the formation and the functioning of these two institutions. The Human Federation's Parliament will elect its Presidency (the President and two Vice Presidents) for a period of 6 years, similar as in the EF. The function of the President will be rotating on a bi-annual basis, when one of the Vice Presidents will become the President. The most important decisions, especially the urgent ones will be the competence of the President. In such situations, he will consult the World Government, the executive body of the Human Federation, on some decisions, which would normally be in his prerogative, if the time allows for that.

Unlike in the EF, the Prime Minister and two deputy Prime Ministers of the World Government will be elected for a period of 6 years by the Parliament rather than by the President. The function of the Prime Minister, similarly as that of the President, will be rotating every two years, when one of the deputy Prime Ministers will become the Prime Minister. There will be a pool of candidates for the Cabinet of 30 ministers and 200 deputy ministers, representing each of the Continents (the numbers are of course only tentative).

Each continent will be allocated a number of candidates for ministers, which will be proportional to its ratio of the world's population. The candidates representing a Continent will be selected randomly and join the pool of candidates for ministers. The actual members of the Cabinet and deputy ministers will be selected by the Prime Minister from the pool of candidates, after the consultation with the President. The whole government will then have to be approved by the Human Federation's Parliament. The President and two Vice Presidents will be the voting members of the World Government. Both the President and the Prime Minister will be accountable to the Parliament. The President will authorize (sign off) all decisions taken by the World Government. He will also represent the Human Federation in any 'negotiations' with Superintelligence. It is quite likely he will be a Transhuman.

Superintelligence in a human-like form will be present at every sitting of the World Government, initially as an advisor. Such an arrangement will most likely run for about a decade from about 2040 to 2050. At this time, the highest-level decisions linked to acting on behalf of the whole Humanity, will probably be made by the Human Federation's President (or Presidency). This is the period when the reformed democracy should be in its adulthood and will be indispensable in keeping the fast-changing civilization in a relative harmony. That is why it is so important that the reforms of democracy do not only respond to past failures but will also be relevant in this crucial future period, when Humanity will pass the baton of the planet's governance to the new species.

CONCLUSIONS

It is worth reminding ourselves that 99.5% of species that have ever lived are now extinct. There were a dozen humanoid species, such as a Neanderthal, which was extinct about 30,000 years ago. Even our own species was almost extinct about 75,000 years ago because of the Toba super volcanic eruption, whose effects lasted 1000 years and left only tens of thousands humans alive. Yes, apocalypses have been a constant theme in the history of the planet Earth.

There is probably little point in discussing natural catastrophic events such as a large asteroid impact, over which we may have no influence at all, at least not now. That is why this book is about a potential man-made (anthropogenic) apocalypse. I started the book by identifying top man-made existential risks and grouping them into three categories. Let me repeat them:

Risks that may become **existential within days** or even hours:

1. Global nuclear war
2. Weaponized AI or cyber soldiers
3. Engineered pandemics and synthetic biology
4. Nanotechnology and experimental technology accident

Risks that may become **existential progressively**:

5. Climate Change over a long time (at least a century)
6. Superintelligence in a short time (a few decades)

Risks that may become **existential because of combinatorial effects** (from days to decades), which I call Global Disorder risks:

7. Global Social Disorder, including global migration as a ‘special’ multi-faceted risk
8. Global Economic Disorder
9. Global Political Disorder

There may also be unknown man-made risks, which may become existential, if combined with other less significant risks such as social disorder. But there is one additional risk, which we must consider. **It is the risk that we may do nothing and just sit and hope for the best.** WWI could have probably been avoided because the real reason for starting it was not the assassination of the Austrian Archduke Franz Ferdinand. Neither, was the cause of the WWII, Poland’s refusal to give Hitler the right to create a corridor through the territory

of Poland leading from German Reich to the free city of Gdansk (Danzig) and Western Prussia.

The real reasons triggering both wars were combinatorial and included the competing forces of rising nationalism, militarism, and expansionism (Germany was not a superpower neither before the WWI or WWII – hence Hitler’s motto – Germany needs ‘Lebensraum’ – a living space). But what finally led to both wars was to a large extent **the indifference among the politicians and populations of France and Britain**, who did not see that the wars far away from their shores (1912-1913 Balkan wars in WWI, and 1938 annexation of Czechoslovakia/the Munich treaty, in WWII). Very soon these wars were fought on their doorsteps. Today, such ‘far away wars’ might be the Russian annexation of Crimea and the invasion of the Ukraine in 2014, contrary to the Budapest Memorandum of 1996, agreed by all superpowers to protect the Ukrainian independence, or the Syrian war. Is this not another sign of indifference to such events that may potentially lead to global catastrophes because they are ‘far away from our shores’?

In this context, perhaps Enrico Fermi, the Italian Nobel laureate in physics was right. His ‘Fermi Paradox’ suggests that the reason there are no civilisations on any of trillions of planets similar to ours, is that “once those civilisations have reached the level of self-destruction, that is precisely what they do – they annihilate themselves”. No wonder then that we have no evidence of their existence. We now have such a capacity. So, why should Humanity be an exception? This is a chilling and unwanted scenario that we all hope will not be fulfilled. But the question is that since we have already reached the level of self-destruction, how can we then avoid it, and become a civilisation that would escape the Fermi Paradox.

Perhaps humans survived so long as the only humanoid species because we have the ability to think generally in positive terms, an essential ingredient to evolutionary survival. Therefore, it does not have to be a utopian view that we can navigate existential risks and ultimately arrive at a world that is safer and more just, where the wealth is distributed much more fairly, and people are generally content.

Well, let’s look then at some positive signs where the world is very slowly getting its acts together. The Ebola crisis has shown that the world can sometimes act very swiftly by abandoning the rule for the sake of urgently creating a viable solution. In this case it was a vaccine developed in just 6 months, rather than in several years if all the rules had been followed. Similarly, the Paris Agreement on Climate Change is not so important because of potential limitation of climate warming-related dangers, but that it happened at all. The fact that 196 countries and organizations committed themselves legally to fulfil

its commitments, which may mean significant potential cost for some of the signatory countries, is very encouraging.

I believe we should now lay down the first paving stone on the road to a new, better world, which will be more sustainable, with a more equal distribution of wealth. Superintelligence, like some other discoveries, e.g. nuclear energy, will change our lives beyond recognition either very positively, or it may simply annihilate us all. Positive outcomes relate mainly to the unprecedented technological capabilities that could significantly improve the quality of our lives and give a new meaning to what our civilisation is about and how it may affect humans as an augmented species. It may also enable the expansion of the human race beyond the solar system.

Today, we are still in charge and can steer the course to that momentous event, so that when Superintelligence emerges, it will become our friend rather than foe. Therefore, we should make real effort right now to minimize the risk that a malicious Superintelligence takes over a total control over us humans. We must remember that the current generation does not only have the responsibility for its own survival but also for the survival of future generations of humans and other species.

My key assumption is that our planet needs to be navigated through a period full of most dangerous, multifaceted existential risks, lasting for about one generation, which can potentially annihilate Humanity. By the end of that period we must be ready to be governed as an almost uniform planetary civilization. To minimize existential risks, we need to carry out significant changes in how democracy and capitalism work and how we share our planet's resources.

So, let me summarize the proposals that I have made throughout the book:

1. To be meaningful, the proposed changes should happen across the globe, in every country and in every nation, but is unrealistic to expect that to happen in a short time
2. Therefore, such changes must be introduced in **most** countries, leaving behind those that will not be ready to join within the next generation. That may create an additional risk of living next to potential adversaries, but the world has no other option
3. Our Planet needs a navigator that will implement such changes; it needs a global organization, which at some stage will be converted into a Human Federation. At the moment there is no such an organization, certainly not the UN, which naturally should have undertaken that role.
4. Such a global organization should take charge of the large part of the world's affairs by about 2030 because it would take at least 10 more years for its decisions to be applied everywhere

5. Additionally, we must complete the task of containing the AI's rapid increase of its capabilities by about 2030, before it will be smarter than humans, even if only in a few domains. Otherwise we may lose control over the goals of the self-learning AI.
6. It is impossible to create a new organization in about 10-15 years, seeing that it took over 20 years to come to the legally binding Paris Agreement on climate change, which is far less complex than the scope of the proposed organization.
7. The only other option is to adapt the existing organization to do that. From the analysis carried out, it looks that the best candidate would be the European Union, if it becomes a multi-zone Federation and gradually takes over the role of the World Government. The other option could be NATO (with additional functionalities) or... China (perhaps a default option to avoid the extinction of all humans).
8. The way the European Union has evolved so far might have led to a Federation in the long term by default. However, in the current circumstances such a slow process is no longer a viable option. A federation is necessary right now for the reasons explained throughout the book.
9. Therefore, the EU needs to federate all countries at once, using perhaps the current initiative (the Future of Europe Conference), if it starts in earnest in the autumn 2020 and if it revises its ultimate objectives
10. Should the Future of Europe Conference change its goals and becomes a Constitutional Convention for the future federation, then such process should use the suggested system of Citizens' Assemblies, which would debate and approve the Constitution, helped by experts throughout the 2-year period
11. However, in the post Covid-19 period, with so many unpredictable factors and chaotic behaviour of many states, the most likely way forward is a partial federalization of the EU. This could be created out of the Eurozone as a whole or formed by a group of a minimum 9 countries, using the current Article 20 of the EU.
12. Whatever the way in which the European Federation is created its fundamental objective should be to fight existential risks. That's what "Federate to Survive!" really means.
13. But since existential threats concern all humans, the European Federation must have a built-in mechanism for a continuous fast enlargement. That can only be achieved by having subsidiary Zones, allowing future members to adapt progressively as they fulfil more stringent criteria of accession to the European Federation state.
14. The future EU Constitution must include a new charter of the Universal Values of Humanity and might incorporate the principles of Consensual Presidential Democracy proposed here (or similar proposals).

I am aware that I have covered many topics from many subject areas to provide a unified view of the world around us, albeit at times they have been simplified

and superficial. Therefore, some errors in interpreting certain facts and drawing conclusions based on incomplete evidence may have been made, for which I apologise. That is the risk of presenting such a wide range of subjects. Moreover, and I have to make this point again, this book is presenting my view and my solutions to the problems that we all face. I would be very happy if just some of the proposals materialize and another few become an inspiration for others to advance this kind of thinking based on Universal Values of Humanity and a deep-down consensual approach to politics.

Whatever thoughts you may have had while reading this book, I would like to encourage you to work hard for your own and others benefit to bring about the world of abundance within the lifetime of most readers. If you are convinced that we need to mobilize all human and material resources to reach such future in just one generation then get active and start thinking long-term, understanding that **pace of change is now exponential**. Have your own vision of a better world and join those who think similarly because only together, united by similar motives of preserving not just our lives, but also the lives of those that come after us, can we achieve that goal.

You can get engaged right now by visiting the website of Sustensis (www.sustensis.co.uk), with its focus on ‘Inspirations for Humanity’s Transition to Coexistence with Superintelligence’.

GLOSSARY

Anthropogenic	Something of man-made origin or caused by man.
Artificial General Intelligence (AGI)	An intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom, and social skills. I use the term Superintelligence in this book rather than AGI.
Artificial Intelligence	An intelligent agent or a machine that surpasses any human being, usually in just one or a few skills, but not all, e.g. playing chess. Quite often it is combined with self-learning capability.
Brexit	Britain's intended exit from the European Union.
Consciousness	Consciousness is probably one of a very few areas in science where the academics cannot agree on the subject they are studying. For the purpose of this book I took the information published in October 2017 by the academics at Collège de France in Paris and would formulate it in layman's terms as follows: "Consciousness is a structural (functional) organization of a physical system, which operates at two levels: 1. Subconsciousness – accepting, storing and retrieving information using huge range of processes with the required algorithms – this is where most human intelligence and knowledge lies. 2 - Actual consciousness containing information about oneself, which it turns into wide range of 'thoughts' all accessible at once to all parts of the brain, which it is able to have continually monitored and processed, outputting them as perceptions and actions." Animals also have consciousness but that is at a much lower level.

Consensual Presidential Democracy	Consensual Presidential Democracy is a system of democracy aimed at governing with maximum consensus, where the voice of the 'losing' minority is always considered. It gives the President exceptionally strong powers against the strongest accountability and recall procedures, to enable him to play a crucial role as a conciliator and moderator between two opposing parties, each represented by one Vice President. This system deepens the separation of legislative and executive powers by having a technocratic government. It also has the widest representation of the electorate, where the representatives to the Parliament are elected using a combined First Past the Post and the Two Rounds System of weighted voting and where a half of the second chamber of the parliament is elected based on a Sortition system.
Constitutional Monarchy	A system, where the monarch is the head of state and a government is elected by conducting free elections.
Direct Democracy	Direct democracy is the type of democracy, in which all eligible citizens can participate in the decision-making process personally, rather than through their representatives.
E-Democracy	The type of democracy, where the voters can exercise their will on the Internet.
European Federation	A proposed name for the federated European Union, proposed to be achieved by 2030.
European Federation Association Area (EFAA)	European Federation Association Area - Zone 4 of the European Federation for members that have individual trade agreements with the European Federation.
European Federation Convergence Area (EFCA)	European Federation Convergence Area - Zone 1 of the European Federation for member states that within a few years will join the European Federation.

European Federation Customs Union (EFCU)	European Federation Customs Union - Zone 3 of the European Federation for countries that are in Customs Union but not in the Single Market.
European Federation Single Market (EFSM)	European Federation Single Market - Zone 2 of the European Federation for countries that are in the Single Market and Customs Union but are not expected to join the European Federation.
Exponential change	This type of change is called exponential, because at each new moment in time (say every year), the value of what we measure (e.g. speed or growth) would double.
GWRF	Global Wealth Redistribution Fund - a fund proposed to be run by the European Federation to lower the wealth inequality world-wide.
Linear change	This type of change is called linear because the growth of, e.g. number of cars sold, will be the same as in the previous period. So, the value of growth is the same in every period.
Nanotechnology	Nanotechnology ("nanotech") is manipulation of matter on an atomic, molecular, and supramolecular scale.
Non-anthropogenic	Something that is not originated by man or not caused by man.
Parliamentary Democracy	A parliamentary system of democratic governance of a state where the government derives its democratic legitimacy through the election of the representatives to the parliament, which in turn selects from its members the Prime Minister and indirectly, the ministers.
Presidential Democracy	A system of governance where the President is the head of state and selects the Prime Minister and sometimes a few key ministers, who are then voted in by the parliament.

Quantum Encryption	Quantum encryption exploits quantum mechanical properties to perform cryptographic tasks in such a way that is impossible to break the key (password) because it would violate the laws of physics.
Quantum Mechanics	Quantum mechanics a fundamental theory in physics, which describes nature at the smallest scales of energy levels of atoms and subatomic particles.
Referendum	A direct voting system, in which an entire electorate is invited to vote on a particular proposal. This may result in the adoption of a new law. In some countries, it is synonymous with a plebiscite or a vote on a ballot question.
Republican Democracy	A Republican system of governance is a version of the Presidential system. The President is the head of state, but the government may fall within a given electoral term and new elections must be called, whereas in the presidential system the same head of state can elect another government (like in France).
Singularity	In the context of Artificial Intelligence, it means Technological Singularity - see below.
Sortition	In governance, sortition means selecting political officials by a random sample from a larger pool of candidates, usually adult who have the right to vote in elections.
Sortition Assembly	This is a one-off Assembly of sortition members selected at random from among the voters to make important political decisions, e.g. to decide on the articles of a constitution.
Sortition Chamber	This is a chamber in the parliament of sortition members selected at random from among the voters to perform the duties identical to Members of Parliament elected through elections.

Superintelligence	An intellect that is much smarter than the best human brains in every field, including scientific creativity, general wisdom, and social skills. In this book it is used instead of the term ‘Artificial General Intelligence’.
Technological Singularity	A point in time when Superintelligence (Artificial General Intelligence), being an intelligent agent smarter than any human being in every aspect of human knowledge, skills, and capabilities, starts re-inventing itself exponentially, through the process of self-learning.
Transpartisan Democracy	A programme of the Danish Party Det Alternativet that focuses on HOW to govern rather than what policies to put in its Manifesto. The WHAT element is a kind of a vague programme, crowd sourced by the party members and aimed at Transition to a sustainable society, supporting entrepreneurship, social entrepreneurship and changing the culture of political dialogue.
Universal Basic Income conditional	A periodic cash payment conditionally delivered by the government to individuals, who meet certain criteria.
Universal Basic Income unconditional	A periodic cash payment unconditionally delivered by the government to individuals without means-testing or work requirement.
Universal Values of Humanity	These are top values of Humanity that apply to humans, animals, and the environment.
Weaponized AI	Artificial Intelligence that can be used as a weapon either in a soft sense - breaking passwords and infiltrating computer systems or in hardware - i.e. AI soldiers.
Weighted Voting System	A system of voting where everybody has a vote, but its weight or value may depend on the knowledge or voter's contributions
World Government	The future organization that would rule Humanity

Bibliography

1. **Wikipedia.** Human evolution. [Online] Wikipedia, 28 02 2018. https://en.wikipedia.org/wiki/Human_evolution.
2. **Marder, Lisa.** What makes us human? [Online] Thought.Co, 04 01 2018. <https://www.thoughtco.com/what-makes-us-human-4150529>.
3. **Williams, Caroline.** *10 Mysteries of you: Blushing.* 5/8/2009, New Scientist.
4. **Bidshahri, Raya.** What Is It That Makes Humans Unique? [Online] 28 12 2017. <https://singularityhub.com/2017/12/28/what-is-it-that-makes-humans-unique/#sm.0000ho5xd6udzf33sh11d0prbs54l>.
5. **Wikipedia.** Kardashev scale. [Online] 2016. https://en.wikipedia.org/wiki/Kardashev_scale.
6. **Kaku, Michio.** 3 Civilization types. [Online] 25 9 2013. <http://www.abovetopsecret.com/forum/thread972919/pg1>.
7. **Maslov, Abraham.** *A theory of human motivation.* 1943, Psychological Review, pp. 50, 370.
8. **Nauert, Rick.** Updated Maslow's Pyramid of Needs. [Online] 30 06 2011. <https://psychcentral.com/news/2010/08/23/updated-maslows-pyramid-of-needs/17144.html>.
9. **Maslow's Pyramid of Human Needs Put to the Test.** [Online] 2015. https://psychcentral.com/news/2011/06/30/maslows-pyramid-of-human-needs-put-to-the-test/27390.html?li_source=LI&li_medium=popular17.
10. **Kurzweil, Ray.** *The Singularity Is Near: When Humans Transcend Biology.* 2006.
11. **Kurzweil, Ray.** *Singularity is Near.* s.l. : Gerald Duckworth & Co Ltd, 9/03/2006.
12. **Teller, Seth** *Ciężkie roboty.* 2013/4, Niezbędnik Inteligenta, p. 94.
13. **Russel Stuart, and Allan Dafoe** *Yes, We Are Worried About the Existential Risk of Artificial Intelligence.* November, 2016, Technology Review, p. 15.

14. **Bostrom, Nick.** *Existential Risk Prevention as Global Priority.* s.l. : Future of Humanity Institute, 2013, pp. 15–3.
15. **Future_of_Humanity_Institute.** Global Catastrophic Risks Survey. [Online] 2008. <https://www.fhi.ox.ac.uk/reports/2008-1.pdf>.
16. **Stern, Nicholas.** *The Economics of Climate Change.* 30/10/2006.
17. **Rees, Martin.** *Our Final Hour: A Scientist Warning.* 2004.
18. **Bostrom, Nick** *Existential Risks - Analyzing Human Extinction Scenarios and Related Hazards.*. 2002, Journal of Evolution and Technology, Vol. 9, No. 1.
19. **Torres, Phil.** *Existential Risks Are More Likely to Kill You Than Terrorism.* 29/6/2016.
20. **Wikipedia.** 1918 flu pandemic. [Online] https://en.wikipedia.org/wiki/1918_flu_pandemic.
21. **Owen Cotton-Barrat, Sebastian Farquha, John Halstead, Stefan Schubert, Andrew Snyder-Beattie.** *Global Catastrophic Risks 2016.* s.l. : Oxford Univeristy Press, 2016.
22. **Global_Challenges_Foundation.** *Global Challenges Report.* 2016.
23. **Bostrom, Nick** *Ethical Issues in Advanced Artificial Intelligence.*. 2002.
24. **Wikipedia.** Geomagnetic reversal. [Online] 10 10 2017. https://en.wikipedia.org/wiki/Geomagnetic_reversal.
25. **Sloan, Rafael Alves Batista and David.** The Conversation. [Online] 14 07 2017. <https://theconversation.com/we-worked-out-what-it-would-take-to-wipe-out-all-life-on-a-planet-and-its-good-news-for-alien-hunters-81006>.
26. **Stout, Martha.** *The Sociopath Next Door.* 8/2/2005.
27. **Torres, Phil** *How likely is an existential catastrophe?* 7/9/2016.
28. **Georgescu-Roegen, Nicholas.** *The Entropy Law and the Economic Process* . s.l. : Cambridge, Massachusetts: Harvard University Press ISBN 0674257804, 1971.
29. **Rees, Martin** *The world in 2050 and beyond.*. 2014, New Statesman.

30. **Chiarelli, B.** *Overpopulation and the Threat of Ecological Disaster: the Need for Global Bioethics* . 1998, *Mankind Quarterly*. 39 (2): 225–230.
31. **Lovgren, Stefan** *Mystery Bee Disappearances Sweeping U.S.*. 2007, *National Geographic News* .
32. **EC_Climate_Change.** Paris Agreement. [Online] 20 12 2015. https://ec.europa.eu/clima/policies/international/negotiations/paris_en.
33. *Why CO2 'Air Capture' Could Be Key to Slowing Global Warming.*
34. **Lackner, Klaus.** 2016, *Yale Environment* 360.
35. **Norcera, Daniel G.** *Harvard Scientist Engineers Bacterium That Inhales CO2, Produces Energy -- A 'Bionic Leaf'*. 29/5/2016, *Financial Times*.
36. **Global_Challenges_Foundation.** *Global Challenges Foundation Report p.20*. Stockholm : s.n., 2017.
37. **Wikipedia.** *Global catastrophic risk*. 2016.
38. **Avner Cohen, Steven Lee.** *Nuclear Weapons and the Future of Humanity: The Fundamental Questions*. p. 237.
39. **Federation_of_American_Scientists.** *Status of World Nuclear Forces*. s.l. : Federation of American Scientists, 28 April 2015.
40. **Shulman, Carl.** *Nuclear winter and human extinction: Q&A with Luke Oman. Overcoming Bias*. . 5 Nov 2012.
41. **Martin, Brian** *Critique of nuclear extinction* . 19 (4): 287–300.. s.l. : *Journal of Peace Research*, 1982.
42. **Physics.** *Atmospheric effects and societal consequences of regional scale nuclear conflicts and acts of individual nuclear terrorism.* *Atmospheric Chemistry and Physics*.
43. **Webber, Philip.** *This Is What Earth Could Look Like After A Nuclear Attack*. [Online] *Huffington Post*, 23 01 2018. http://www.huffingtonpost.co.uk/entry/this-is-what-earth-could-look-like-after-a-nuclear-attack_uk_5a660e34e4b0022830050df1.
44. **Keck, Zachary.** *Billions Could Die If India and Pakistan Start a Nuclear War*. [Online] 21 07 2017. <http://nationalinterest.org/blog/the-buzz/billions-could-die-if-india-pakistan-start-nuclear-war-21623> .

44. **Institute for Science and International Security.** Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant? [Online] 22 12 2010. http://isis-online.org/uploads/isis-reports/documents/stuxnet_FEP_22Dec2010.pdf.
45. **Kalman, Aaron.** Only ‘the nuclear option’ can work against Iran, former IDF chief says. [Online] 10 11 2012. <https://www.timesofisrael.com/only-the-nuclear-option-can-work-against-iran-former-idf-chief-says/>.
46. **Revesz, Rachael.** North Korea EMP attack could ‘shut down US power grid and kill 90% of Americans’. [Online] Independent, 17 10 2017. <http://www.independent.co.uk/news/world/americas/north-korea-us-attack-emp-power-grid-kill-90-per-cent-american-population-electromagnetic-pulse-a8002756.html>.
47. **Isabel Reynolds, Enda Curran.** What nuclear war between the US and North Korea might look like. [Online] Independent, 09 08 2017. <http://www.independent.co.uk/news/world/americas/us-north-korea-donald-trump-kim-jong-un-nuclear-missiles-military-war-explainer-pyongyang-guam-a7883581.html>.
48. **Williams, Martin.** FactCheck Q&A: Could there be a nuclear Armageddon? [Online] 04 07 2017. <https://www.channel4.com/news/factcheck/factcheck-qa-will-there-be-a-nuclear-armageddon>.
49. **Hoover Institution.** Next Steps in Reducing Nuclear Risks. [Online] Hoover Institution, 06 03 2013. <https://www.hoover.org/research/next-steps-reducing-nuclear-risks>.
50. **Union of Concerned Scientists.** Reducing the Risk of Nuclear War. [Online] Union of Concerned Scientists. https://www.ucsusa.org/nuclear-weapons/us-nuclear-weapons-policy/reducing-the-risk#.Wo_eA6hl-M8.
51. **Frey, Thomas.** Weaponized A.I. – 36 Early Examples. [Online] Futurist Speaker, 08 08 2017. <http://www.futuristspeaker.com/business-trends/weaponized-a-i-36-early-examples/>.
52. **Santos, Paulo E.** Autonomous weapons and the curse of history. [Online] 23 11 2015. <https://thebulletin.org/autonomous-weapons->

civilian-safety-and-regulation-versus-prohibition/banning-and-regulating-autonomous-weapons.

53. **Drexler, Eric.** Productive Nanosystem: From molecules to superproducts. [Online] 14/3/2006.

<https://www.youtube.com/watch?v=vEYN18d7gHg>.

54. **Phoenix, Chris and Treder, Mike.** *Chapter 21: Nanotechnology as global catastrophic risk.* In *Bostrom, Nick; Cirkovic, Milan M. Global catastrophic risks.* s.l. : Oxford University Press, 2008.

55. *Dangers of Molecular Manufacturing.*

Centre_for_Responsible_Nanotechnology. s.l. : Centre for Responsible Nanotechnology, 19 July 2014.

56. **SA, Frank.** *Models of parasite virulence* *Q Rev Biol.* 71. March 1996, *Q Rev Biol.* 71 .

57. **Noun, Ali and Chyba, Christopher F.** *Chapter 20: Biotechnology and biosecurity,* in *Bostrom, Nick; Cirkovic, Milan M. Global Catastrophic Risks.* s.l. : Oxford University Press, 2008.

58. **Wikipedia.** Wikipedia. [Online] *Anthropological Quarterly*, 10 12 2017. https://en.wikipedia.org/wiki/Anthropological_Quarterly.

59. **Dolan, Kerry A.** *Forbes 2017 Billionaires List: Meet The Richest People On The Planet.* [Online] *Forbes*, 20 03 2017.

<https://www.forbes.com/sites/kerryadolan/2017/03/20/forbes-2017-billionaires-list-meet-the-richest-people-on-the-planet/#3aa7df7662ff>.

60. **McMinn, David.** *PLANNED OBSOLESCENCE: THE ULTIMATE ECONOMIC INEFFICIENCY.* [Online] David McMinn, 2017. <http://www.davidmcminn.com/ngc/pages/obsol.htm>.

61. **Third World Network.** *Why tax havens must go!* [Online] *Third World Network.*

http://www.cadtm.org/spip.php?page=imprimer&id_article=13847.

62. **Trompenaaras, Charles Hampden-Turner and Fons.** *Nine Visions of Capitalisms - Chapter 7.* 2015.

63. **Bostrom_Nick.** *Existential Risks.* 2002.

64. **Wikipedia.** *Transnationalism.* [Online] *Wikipedia*, 30 01 2018. <https://en.wikipedia.org/wiki/Transnationalism>.

65. **Australian National University.** *The Aliens Are Silent Because They Are Extinct.* s.l. : Australian National University, 21/1/2016.
66. **Frank, Adam.** Is a Climate Disaster Inevitable? . *The New York Times.* 17/1/2015.
67. **Cohen, Hsin-Yi.** What is Intelligence and How is it Measured? [Online] 10 02 2016. <http://www.aboutintelligence.co.uk/what-intelligence.html>.
68. **Wikipedia.** Technological singularity. [Online] Wikipedia, 01 03 2018.
https://en.wikipedia.org/wiki/Technological_singularity#cite_note-Carvalko.2C_Joseph_2012-3.
69. **Urban, Tim.** The AI Revolution: The road to Superintelligence. [Online] 25 01 2015. <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>.
70. **Wikipedia.** Intelligence Quotient. [Online]
https://en.wikipedia.org/wiki/Intelligence_quotient.
71. **Feng Liu, Yong Shi, and Ying Liu.** Google AI vs Siri vs Bing - IQ tests show one is smartest by a mile. [Online] 03 10 2017.
<http://techgroundnews.com/apple/google-ai-vs-siri-vs-bing-iq-tests-show-one-is-smartest-by-a-mile/>.
72. **Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang and Owain Evans.** *When Will AI Exceed Human Performance? Evidence from AI Experts.* Oxford : Future of Humanity Institute, 30/05/2017.
73. **Reedy, Dom Galeon and Christianna.** Kurzweil Claims That the Singularity Will Happen by 2045. [Online] 05 10 2017.
<https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045/>.
74. **Green, Tristan.** Google's AI can create better machine-learning code than the researchers who made it. [Online] 17 10 2017.
<https://thenextweb.com/artificial-intelligence/2017/10/16/googles-ai-can-create-better-machine-learning-code-than-the-researchers-who-made-it/>.
75. **Rejcek, Peter.** AI Uses Titan Supercomputer to Create Deep Neural Nets in Less Than a Day. [Online] Singularity Hub, 03 01 2018.

https://singularityhub.com/2018/01/03/ai-uses-titan-supercomputer-to-create-deep-neural-nets-in-less-than-a-day/?utm_source=Singularity+Hub+Newsletter&utm_campaign=e481a6ee4d-

[Hub_Weekly_Newsletter&utm_medium=email&utm_term=0_f0cf60cd-ae-e481a6ee4d-58229305.](https://singularityhub.com/2018/01/03/ai-uses-titan-supercomputer-to-create-deep-neural-nets-in-less-than-a-day/?utm_source=Singularity+Hub+Newsletter&utm_campaign=e481a6ee4d-58229305)

76. **Kurywczak, Eugene.** The Dynamic Eternal Universe. [Online] 15 05 2014.

[https://books.google.co.uk/books?id=b96ZAwAAQBAJ&pg=PA15&lp_g=PA15&dq=nothingness+is+eternal&source=bl&ots=Dv7a7BvQja&sig=xoi7wIbTNY5y4Z5WsVvCPYbKO5I&hl=en&sa=X&ved=0ahUKEwijn7MzOvubZAhUMCsAKHcxqDPM4ChDoAQhAMAQ#v=onepage&q=nothingness%20is%20eternal&f=false.](https://books.google.co.uk/books?id=b96ZAwAAQBAJ&pg=PA15&lp_g=PA15&dq=nothingness+is+eternal&source=bl&ots=Dv7a7BvQja&sig=xoi7wIbTNY5y4Z5WsVvCPYbKO5I&hl=en&sa=X&ved=0ahUKEwijn7MzOvubZAhUMCsAKHcxqDPM4ChDoAQhAMAQ#v=onepage&q=nothingness%20is%20eternal&f=false)

77. **Bostrom, Nick.** *Superintelligence: Paths, Dangers, Strategies.* 2014.

78. **Yampolskiy, Roman V.** Fighting malevolent AI: artificial intelligence, meet cybersecurity. [Online] 13 06 2016.

[http://theconversation.com/fighting-malevolent-ai-artificial-intelligence-meet-cybersecurity-60361.](http://theconversation.com/fighting-malevolent-ai-artificial-intelligence-meet-cybersecurity-60361)

79. **Torres, Phil.** Top three strategies for avoiding existential risks. [Online] Institute for Ethics and Emerging Technologies, 13 01 2016.

[https://ieet.org/index.php/IEET2/more/torres20120213.](https://ieet.org/index.php/IEET2/more/torres20120213)

80. **Müller, Vincent C. and Bostrom, Nick.** *Future Progress in Artificial Intelligence: A Survey of expert Opinion.* 2014.

81. **BBC.** Massive ransomware infection hits computers in 99 countries.

[Online] 13 05 2017. [http://www.bbc.co.uk/news/technology-39901382.](http://www.bbc.co.uk/news/technology-39901382)

82. **Mali, Malhar.** *How the World Will End; Nuclear Armageddon, A.I., Climate Change - Interview with Phil Thores of X-Risks.* 12 10 2016.

83. **Domonoske, Camila.** Elon Musk Warns Governors: Artificial Intelligence Poses 'Existential Risk'. [Online] 17 July 2017.

[https://www.npr.org/sections/thetwo-way/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk.](https://www.npr.org/sections/thetwo-way/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk)

84. **Assimov, Isaac.** *I, Robot*. 1950.
85. **Danaher, John.** Bostrom on Superintelligence (5): Limiting an AI's Capabilities . [Online] 09 08 2014.
<http://philosophicaldisquisitions.blogspot.co.uk/2014/08/bostrom-on-superintelligence-5-limiting.html>.
86. **Czarnecki, Tony.** 'Who could save Humanity from Superintelligence? s.l. : Amazon, 2018.
87. **Tegmark, Max.** *Life 3.0: Being Human in the Age of Artificial Intelligence*. s.l. : Amazon, 2017.
88. **Cookson, Clive.** Superintelligence: a space odyssey. [Online] 30 08 2017. <https://www.ft.com/content/31176c28-8bea-11e7-9084-d0c17942ba93>.
89. **Moody, Tom Whipple and Oliver.** Stephen Hawking on humanity . [Online] The Times, 07 03 2017.
<https://www.thetimes.co.uk/edition/news/hawking-on-humanity-and-corbbyn-jk88zx0w2>.
90. **Rees, Martin.** Martin Rees: The world in 2050 and beyond. [Online] 26 11 2014. <https://www.newstatesman.com/sci-tech/2014/11/martin-rees-world-2050-and-beyond>.
91. **Policy, Institute for Global.** World Federalist Movement. [Online] 2016. <http://www.wfm-igp.org/>.
92. **Politico.** Europeans are more positive about future of the EU — except the Brits. [Online] Politico.
<https://www.politico.eu/article/europeans-are-more-positive-about-future-of-the-eu-except-the-brits/>.
93. **Stone, Jon.** More Europeans than ever say they feel like citizens of the EU. [Online] The Independent, 02 08 2017.
<http://www.independent.co.uk/news/uk/politics/eu-brexiteuropean-union-citizens-feel-like-eurobarometer-survey-results-a7872916.html>.
94. **Institute of Fiscal Studies.** Is our tax system fair? It depends... [Online] 03 11 2017. <https://www.ifs.org.uk/publications/10038>.
95. **OECD.** GDP long-term forecast. [Online] OECD, 2015.
<https://data.oecd.org/gdp/gdp-long-term-forecast.htm>.

96. **PWC.** The World in 2050. [Online] PWC, 15 02 2015.
<https://www.pwc.com/gx/en/issues/the-economy/assets/world-in-2050-february-2015.pdf>.
97. **Williams, Matt.** FALCON HEAVY VS. SATURN V. [Online] Universe Today, 16 01 2018.
<https://www.universetoday.com/129989/saturn-v-vs-falcon-heavy/>.
98. **Wang, Brian.** Lab grown meat prices have dropped 30,000 times in less than four years and are about 3-4 times more expensive than regular ground beef. [Online] Next Big Future, 19 02 2017.
<https://www.nextbigfuture.com/2017/02/lab-grown-meat-prices-have-dropped.html>.
99. **Collinson, Patrick.** Finland is the happiest country in the world, says UN report. [Online] The Guardian, 14 03 2018.
<https://www.theguardian.com/world/2018/mar/14/finland-happiest-country-world-un-report>.
100. **Diamandis, Peter.** Why the Cost of Living Is Poised to Plummet in the Next 20 Years. [Online] Singularity Hub, 18 07 2016.
<https://singularityhub.com/2016/07/18/why-the-cost-of-living-is-poised-to-plummet-in-the-next-20-years/>.
101. **Institute, Wellcome Sanger.** Whole genome sequencing will ‘transform the research landscape for a wide range of diseases’. *Wellcome Sanger Institute*. [Online] Wellcome Sanger Institute, 08 04 2018. <http://www.sanger.ac.uk/news/view/whole-genome-sequencing-will-transform-research-landscape-wide-range-diseases>.
102. **World Atlas.** The Continents Of The World Per Capita GDP. [Online] World Atlas, 01 08 2017.
<https://www.worldatlas.com/articles/the-continents-of-the-world-by-gdp-per-capita.html>.
103. **World Atlas.** The Poorest Countries In The World. [Online] World Atlas, 06 12 2017. <https://www.worldatlas.com/articles/the-poorest-countries-in-the-world.html>.
104. **Forum, Namibia NGO.** *Basic Income Grant Coalition*. 2008.