THE FRONTIERS COLLECTION

Anthony Aguirre
Brendan Foster
Zeeya Merali (Eds.)

# HOW SHOULD HUMANITY STEER THE FUTURE?

FQXi
FOUNDATIONAL QUESTIONS INSTITUTE

Springer

# THE FRONTIERS COLLECTION

**Series editors**

Avshalom C. Elitzur
Unit of Interdisciplinary Studies, Bar-Ilan University, 52900 Ramat-Gan, Israel
e-mail: avshalom.elitzur@weizmann.ac.il

Laura Mersini-Houghton
Department of Physics, University of North Carolina, Chapel Hill,
NC 27599-3255, USA
e-mail: mersini@physics.unc.edu

T. Padmanabhan
Inter University Centre for Astronomy and Astrophysics (IUCAA), Pune, India

Maximilian Schlosshauer
Department of Physics, University of Portland, Portland, OR 97203, USA
e-mail: schlossh@up.edu

Mark P. Silverman
Department of Physics, Trinity College, Hartford, CT 06106, USA
e-mail: mark.silverman@trincoll.edu

Jack A. Tuszynski
Department of Physics, University of Alberta, Edmonton, AB T6G 1Z2, Canada
e-mail: jtus@phys.ualberta.ca

Rüdiger Vaas
Center for Philosophy and Foundations of Science, University of Giessen,
35394 Giessen, Germany
e-mail: ruediger.vaas@t-online.de

**THE FRONTIERS COLLECTION**

*Series Editors*

A.C. Elitzur L. Mersini-Houghton T. Padmanabhan M. Schlosshauer
M.P. Silverman J.A. Tuszynski R. Vaas

The books in this collection are devoted to challenging and open problems at the forefront of modern science, including related philosophical debates. In contrast to typical research monographs, however, they strive to present their topics in a manner accessible also to scientifically literate non-specialists wishing to gain insight into the deeper implications and fascinating questions involved. Taken as a whole, the series reflects the need for a fundamental and interdisciplinary approach to modern science. Furthermore, it is intended to encourage active scientists in all areas to ponder over important and perhaps controversial issues beyond their own speciality. Extending from quantum physics and relativity to entropy, consciousness and complex systems—the Frontiers Collection will inspire readers to push back the frontiers of their own knowledge.

More information about this series at http://www.springer.com/series/5342

For a full list of published titles, please see back of book or springer.com/series/5342

Anthony Aguirre · Brendan Foster
Zeeya Merali
Editors

# HOW SHOULD HUMANITY STEER THE FUTURE?

FQXi
Foundational Questions Institute

Springer

*Editors*
Anthony Aguirre
Department of Physics
University of California
Santa Cruz, CA
USA

Brendan Foster
Foundational Questions Institute
New York, NY
USA

Zeeya Merali
Foundational Questions Institute
New York, NY
USA

ISSN 1612-3018 ISSN 2197-6619 (electronic)
THE FRONTIERS COLLECTION
ISBN 978-3-319-20716-2 ISBN 978-3-319-20717-9 (eBook)
DOI 10.1007/978-3-319-20717-9

Library of Congress Control Number: 2015947398

Springer Cham Heidelberg New York Dordrecht London


Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

# Preface

This book is a collaborative project between Springer and The Foundational Questions Institute (FQXi). In keeping with both the tradition of Springer's Frontiers Collection and the mission of FQXi, it provides stimulating insights into a frontier area of science, while remaining accessible enough to benefit a non-specialist audience.

FQXi is an independent, nonprofit organization that was founded in 2006. It aims to catalyze, support, and disseminate research on questions at the foundations of physics and cosmology.

The central aim of FQXi is to fund and inspire research and innovation that is integral to a deep understanding of reality, but which may not be readily supported by conventional funding sources. Historically, physics and cosmology have offered a scientific framework for comprehending the core of reality. Many giants of modern science—such as Einstein, Bohr, Schrödinger, and Heisenberg—were also passionately concerned with, and inspired by, deep philosophical nuances of the novel notions of reality they were exploring. Yet, such questions are often overlooked by traditional funding agencies.

Often, grant-making and research organizations institutionalize a pragmatic approach, primarily funding incremental investigations that use known methods and familiar conceptual frameworks, rather than the uncertain and often interdisciplinary methods required to develop and comprehend prospective revolutions in physics and cosmology. As a result, even eminent scientists can struggle to secure funding for some of the questions they find most engaging, while younger thinkers find little support, freedom, or career possibilities unless they hew to such strictures.

FQXi views foundational questions not as pointless speculation or misguided effort, but as critical and essential inquiry of relevance to us all. The Institute is dedicated to redressing these shortcomings by creating a vibrant, worldwide community of scientists, top thinkers and outreach specialists who tackle deep questions in physics, cosmology, and related fields. FQXi is also committed to engaging with the public and communicating the implications of this foundational research for the growth of human understanding.

As part of this endeavor, FQXi organizes an annual essay contest, which is open to everyone, from professional researchers to members of the public. These contests are designed to focus minds and efforts on deep questions that could have a profound impact across multiple disciplines. The contest winners are chosen by a combination of input from entrants, FQXi Members, and a panel of judges, and up to twenty prizes are awarded. Each year, the contest features well over a hundred entries, stimulating ongoing online discussion long after the close of the contest.

We are delighted to share this collection, inspired by the 2014 contest, "How Should Humanity Steer the Future?" In line with our desire to bring foundational questions to the widest possible audience, the entries, in their original form, were written in a style that was suitable for the general public. In this book, which is aimed at an interdisciplinary scientific audience, the authors have been invited to expand upon their original essays and include technical details and discussion that may enhance their essays for a more professional readership, while remaining accessible to nonspecialists in their field.

FQXi would like to thank our contest partners: Jaan Tallin, The Gruber Foundation, The John Templeton Foundation, and Scientific American. The editors are indebted to FQXi's scientific director, Max Tegmark, and managing director, Kavita Rajanna, who were instrumental in the development of the contest. We are also grateful to Angela Lahee at Springer for her guidance and support in driving this project forward.

2015

Anthony Aguirre
Brendan Foster
Zeeya Merali
www.fqxi.org

# Contents

# Chapter 1
# Introduction

**Anthony Aguirre, Brendan Foster and Zeeya Merali**

*Science fiction writers foresee the inevitable, and although problems and catastrophes may be inevitable, solutions are not.*
Isaac Asimov (1931) [1]

*We are in danger of destroying ourselves by our greed and stupidity. We cannot remain looking inwards at ourselves on a small and increasingly polluted and overcrowded planet.*
Stephen Hawking (2010) [2]

Improving the future for our civilization is one of the foremost goals of both the sciences and the humanities. These endeavours allow us to learn from both our past mistakes and successes, to anticipate potential catastrophes, and to develop technologies and lines of thinking to preempt them. Yet dystopic visions of the future—often based on the unchecked rise of the very scientific and technological innovations designed to help society—abound in literature and film, while optimistic ones are more rare.

In 2014, FQXi launched an essay contest with the aim of redressing the balance by encouraging entrants to think about ways to avoid potentially self-fulfilling prophecies of doom and gloom. "How," we asked, "should humanity steer the future?"

This was one of the broadest questions that we had yet posed for an essay contest, and required participants to not only imagine future pitfalls, but also to outline

A. Aguirre (✉)
Department of Physics, University of California, Santa Cruz, CA, USA
e-mail: aguirre@scipp.ucsc.edu

B. Foster · Z. Merali
Foundational Questions Institute, New York, NY, USA
e-mail: foster@fqxi.org

Z. Merali
e-mail: merali@fqxi.org



A. Aguirre et al. (eds.), *How Should Humanity Steer the Future?*,
The Frontiers Collection, DOI 10.1007/978-3-319-20717-9_1

practical strategies to mitigate them. Our ever-deepening understanding of physics has enabled technologies and ways of thinking about our place in the world that have dramatically transformed humanity, and the world that we live in, over the past several hundred years. Some of the resulting problems that will face future generations are already apparent. It will require global efforts to address human-induced climate change, for instance. Yet, as we have seen, it is often difficult to persuade governments and the public to establish policies and habits now that may only reap benefits over the longterm. Other threats to humanity that could arise from future technology, such as artificial intelligence, have barely even entered serious public discussion. Many others will take an unknown form that we have yet to imagine based on the radically different modes of thought and fundamentally new technologies that could become relevant in the coming decades.

In this vein, we asked participants to consider what they believe the best state that humanity could realistically achieve might be, what plan would be needed to reach that point, and who would need to implement that plan. The contest drew 155 entries from thinkers both within and outside the academic system. It proved a resounding success, raising many new lines of inquiry and demonstrating the same creativity, big-picture thinking and depth of understanding seen in previous essay contests. This success, and the urgency of many of the issues brought to light, inspired the foundation of a separate body, the Future of Life Institute (http://futureoflife.org/), which supports initiatives for safeguarding life and developing optimistic visions of the future. FLI has subsequently grown rapidly, with the successful launch of several initiatives addressing the future promise and perils of artificial intelligence.

This volume brings together the top 14 prize-winning entries from the contest. Some identify particular risks to humanity's security, while others propose general changes that could be made now to education and research in order to arm society against threats of any form—whether natural, human-induced, or even from alien civilizations. Still others address how to make society receptive to any proposed changes.

Our first-prize winner, Sabine Hossenfelder, challenges the value of the essay question itself. In Chap. 2, she notes that even if we knew how best to steer the future of humanity, that knowledge will be of little use if the wider population does not enforce it. She sketches a strategy for disseminating insights to the public in a palatable manner to maximise their impact, enabling people to evaluate possible courses of action for themselves in an informed manner.

The next two chapters deal with assessing the specific form of longterm risks. Given how difficult it is to predict tomorrow's weather, in Chap. 3, Tommaso Bolognesi, considers how best to accurately simulate far-future scenarios for humanity's fate based on existing data sets. In Chap. 4, Daniel Dewey calls for governments to invest in research into possible threats arising from biological engineering and artificial intelligence.

Chapters 5–8 offer strategies to arm humanity against catastrophes, whatever they may be, by changing people's attitudes about their influence over the future. Preston Estep III and Alexander Hoekstra, in Chap. 5, advocate focusing on techniques for strengthening the human mind. Dean Rickles argues that people often underestimate

their ability to affect the future. In Chap. 6, he suggests that inspiration to rectify this could come from interpretations of quantum mechanics that highlight the role of the observer on measurements and from the philosophy of time. Rick Searle similarly discusses how the impact of technology on society has led to a lost sense of freedom over the future. In Chap. 7, he argues that this could be remedied by re-establishing the "Utopian ideal". By contrast, Tejinder Singh makes the case that fixating on the past or obsessing anxiously over the future can have a negative effect on the mind. Instead, he advises that humanity should learn to live in the here and now. Enlightenment, he says in Chap. 8, is not for the Buddha alone.

A number of winners proposed ways to improve current science education and research. In Chap. 9, Travis Norsen argues that science teaching should be less dogmatic, with more emphasis on the historical development of ideas and scientific controversies, so that scientists are better equipped to deal with contentious issues in the future. Jonathan Dickau values the role of play in learning and physics research and, in Chap. 10, he argues that recognising this will fuel innovation.

In Chap. 11, Mohammed Khalil proposes a number of changes within the academic system, including developing new specialisations at undergraduate level to deal specifically with energy solutions, encouraging collaboration between various disciplines, enhancing public understanding of science through online courses, and using *Wikipedia* as a model for generating online review articles summarising new research. The issue of how to store information over the longterm is also addressed, in Chap. 12, by Jens Niemeyer, who notes that as ever-increasing amounts of data are held in digital form, the risk of losing vast tracts of knowledge in a global disaster is also raised. He argues that a secure physical repository is needed to protect humanity's heritage.

Chapters 13–15 look beyond Earth when considering global security. In Chap. 13, George Gantz invites people to draw on humanity's most positive values such as love, respect, and humility, to prepare them for possible first contact with an alien civilization. Flavio Mercati uses lessons from history to make the case that humanity needs to achieve equilibrium with its environment. In Chap. 14, he considers future scenarios in which people will need to terraform and colonise other planets and argues that preserving biodiversity will be essential for their success. Chapter 15 closes the volume with a novel work of fiction by Georgina Parry that imagines a highly technological world in the wake of overpopulation and climate change. She explores the issues surrounding the society of survivors as they contemplate migration to another world.

This compilation brings together the most diverse range of winners of any of FQXi's essay contests. The contributors to this volume include academic researchers from the fields of high energy physics, theoretical and computational cosmology, philosophy and quantum gravity, and those who now work, or have worked previously, in genetics, aspects of mathematical modelling, software engineering, audio and video production, business, and science education. This mix is appropriate given the broad scope of the question that FQXi posed and its widescale potential impact. It serves to highlight the importance of interdisciplinary collaboration when considering the longterm future of our civilization.

## References

1. Asimov, I.: How Easy to See the Future, Natural History magazine (April 1975); later published in Asimov on Science Fiction (1981)
2. Hawking, S.: Interview with Larry King, CNN, 10 September 2010

# Chapter 2
# How to Save the World

## In Five Simple Steps

**Sabine Hossenfelder**

**Abstract** If you knew how humanity should steer the future, what difference would it make? The major challenge that humanity faces today is not that we lack ideas for what to do, as I am sure this essay contest will document. No, the major challenge, the mother of all problems, is to convert these ideas into courses of action. We fail to act in the face of global problems because we do not have an intuitive grasp on the consequences of collective human behavior, are prone to cognitive biases, and easily overwhelmed by data. We are also lazy and if intuition fails us, inertia takes over. How many people will read these brilliant essays? For the individual, evaluating possible courses of action to address interrelated problems in highly connected social, economic and ecological networks is presently too costly. The necessary information may exist, even be accessible, but it is too expensive in terms of time and energy. To steer the future, information about our dynamical and multi-layered networks has to become cheap and almost effortless to use. Only then, when we can make informed decisions by feeling rather than thinking, will we be able to act and respond to the challenges we face.

### 2.1 The Problem

The most remarkable fact about humans is the utter uselessness of our infants. Humans, in contrast to all other species, must learn almost everything necessary for survival. It takes us a long time to reach maturity, time in which parents have to prevent their offspring from eating sand, chopping off fingers, or accidentally wiping out the human race by growing super-resistant bugs behind their ears.

But our ability to learn, combined with technics to communicate information, is also what enables us to adapt to changing environments faster than gene selection could possibly achieve. We are awed by inborn knowledge—butterflies that recall routes of their ancestors—but we outpaced our competitors by changing the rules of what it means "to adapt" itself. We do not wait for physiological changes to result

S. Hossenfelder (✉)
Nordita, Roslagstullsbacken 23, 10691 Stockholm, Sweden
e-mail: sabine.hossenfelder@gmail.com



A. Aguirre et al. (eds.), *How Should Humanity Steer the Future?*,
The Frontiers Collection, DOI 10.1007/978-3-319-20717-9_2

in better Darwinian fitness. Instead, we modify our environment, ourselves, and our interaction with the environment to get there faster, to "fit" better than any other species.

The root of the problems that humanity faces today is that our adaptation as a species has fallen behind the changes we have induced ourselves. As human interactions get more complex, as networks spread globally and become tightly coupled, we need systems that are able to learn and in return help let us learn about the system. But we don't have them.

The political, economic, and social systems that govern our lives are presently adaptive by trial and error. But much like gene selection is too slow to have yet adapted humans to a mostly sedentary city life and goat memes, the adaptation of our systems by trial and error is too slow to solve the problems we presently face. May that be climate change, the global water crisis, the big garbage swirl, or the fragility of our financial systems—our inability to process these problems means that we, as the actors in the system, do not respond, indeed cannot respond, to the information we have in a timely manner. And so, problems persist and build up.

The necessary information for individuals to learn and react to systemic trends may be available, even accessible, but it is too expensive. Information is presently costly, not necessarily financially, but in the amount of effort required to obtain and understand it. Relevant information is too difficult to find or comprehend and doing so requires too much time and energy. Blaming people for being politically disinterested, scientifically illiterate, or plainly unintellectual doesn't do anything to address the costliness of information and thus doesn't do justice to the origin of the problem. The individual investment necessary to process information about trends and relations in our systems is currently too high and personal benefits do not outweigh the disadvantages.

There is no shortage of well-intentioned institutions and organizations that aim at one or several of humanity's problems. The biggest, most existential, problems have been collected by the Future of Humanity Institute [1]. These are the problems that can lead to extinction, near-extinction or progress stagnation of the human species, including but not limited to nuclear accidents, asteroid impacts, artificial intelligence and biotechnology. In case you're not a worrier, check their webpage.

The Future of Humanity Institute is also a point in case because it has next to no influence on the future of humanity. The same can be said about all other initiatives that collect data about our global networks and use buzzwords like complexity and interdisciplinarity. For example the United Nations Global Pulse [2], the FutureICT [3] or the PSIR Model [4]. All such initiatives fail to address the main problem, which isn't to collect information, but feeding this information back into the system, back to the many humans who are the initiators of change.

Some believe quantum computing will solve our problems. It won't. Computation alone cannot solve our problems for the same reason political utopias, however beautiful and ingenious, never solved any problems: Because humans don't care what somebody or some thing thinks they should be doing. They'll do whatever they please. The only way to change their ways is to please them. Please them differently than before, and change will follow.

No, the biggest challenge mankind faces today is not the development of some breakthrough technology. The biggest challenge is to create a society whose institutions integrate the knowledge that must precede any breakthrough technology: The knowledge about the systems themselves that is necessary for the realization, adaptation and use of technologies. All of our big problems today speak of our failure, not to envision solutions, but to turn our ideas and knowledge into reality. We have a social problem, not a technological one.

We reached this gridlock because the human brain did not evolve to understand the consequences of individual actions in networks of billions of people. We are bad in making good long-term decisions and do not care much what happens in other parts of the planet to people we have not and will most likely never meet. We have no intuitive grasp on the collective behavior of large groups and their impact on our environment, and what little grasp we have is prone to cognitive biases and statistical errors, many of which are now subject of new scientific areas like game theory, behavioral economics and decision science.

These cognitive shortcomings are not only obstacles to solving our problems, they *are* the problem. But these are obstacles that science can overcome.

## 2.2 The Solution

The human brain has the capacity to evaluate decisions that have long-term and large-scale consequences. However, frequently decisions which are beneficial on long time- or distance scales conflict with those on short time- or distance scales. Due to evolutionary developed reward circuits, this conflict is often resolved in favor of short times and distances.[1]

But we know how to solve these problems. We solve them by bringing close that what is far away. This is why people in weight-loss programs (distant) are encouraged to reward themselves (close) for holding onto their diet. This is why they pin photos (distant) on their fridge (close). This is why the World Wildlife Fund lets you adopt baby animals of endangered species (distant) and sends you a certificate (close). This is why you are shown all the photos of hungry, ill, injured or otherwise suffering children. You get the picture. It brings distant information closer and taps onto your emotional responses, which is a fast, simple, and effective reaction. It wires back into the circuits that your brain is used to work with. This wiring can be abused, all right. But used the right way, it carries the solution to our problem.

"Gamification" is a recent variant of this mechanism. Gamification is growing popular to help people balance their own priorities, typically by providing instant rewards (in terms of collecting points) for behavior users previously themselves identified as desirable (say, eating healthy). Seen from a system's perspective, this is an external feedback loop that allows humans to use old brain circuits to adapt good

[1]Here, with "distance" I am referring not necessarily to spatial distance, but to distance in social networks and other infrastructure networks.

(here: healthy) behavior faster than gene selection could achieve with a turnover rate of many generations. The interesting aspect about gamification is how little is necessary to make this feedback loop work. All it takes is a simple and intuitive visualization that lets users immediately grasp how well an action matches with their stated goals. The keywords here are: Simple, intuitive, and immediate. This is cheap information.

The solution to our problems is a generalization of this feedback loop: To give people access to cheap information about the consequences of collective human actions, and in return use their reaction to this information to improve the system, i.e. the way individual actions are coordinated.

The point here is not to manipulate people into changing their ways because I or you or some supercomputer thinks it would be better if we'd do more of this or more of that. The point is to help people make decisions. The way we presently make decisions, part of our priorities remain neglected because we cannot assess how well we would be working towards them. It's too complicated, too costly. But it's not like we are happy with this. Most people notice the tension, the neglect of some of their priorities, and are left with bad consciousness, the nagging voice that says you should make better decisions. If only you had the time and it wasn't so difficult.

The feedback system that we need has to give the user an intuitive feeling for how well a decision matches with recorded priorities. If such a feedback in the future can be given by a brain implant, it will be like an additional sense. How does this decision taste? How well does it match with my preferences? Does this choice look harmonious? Does it sound good? Such a feedback is the natural extension of our ability to judge the result of our actions in small groups. This is what it takes to make information cheap, really cheap, so that using it becomes almost effortless.

This feedback loop might include for example information about how well buying a product matches with the relevance one has assigned to certain health goals or its environmental impact or its contribution to the local economy. This is information which a customer doesn't normally have when making a purchase (though economic theory maintains it is taken into account). And even if they had the information, they probably wouldn't study it.

Other examples are questions like: If I dispose of that plastic bottle here, how likely is it to be recycled or to end up in the ocean? If I buy the fair trade coffee, does it work towards something I value? Do I help the homeless guy more by giving him some dollars or by donating that money somewhere? How much of the tax I pay on this item subsidizes projects I support? It's not that people do not care. It's just that in practice it takes too much effort to look into the details. And they actually do not want to know the details. All they want to know is whether, according to best present knowledge, a certain decision works towards their goals. And most of the time that is really all they need to know.

Let me use another example, a somewhat shocking one that however illustrates well distance among people. A recent study by researchers from Princeton University asked participants to judge the competence of political candidates by split-second looks at photos. It turned out that this snap judgment predicted very well who would eventually be voted [5].

How incredibly shallow we are. But forgive us. We decode human faces constantly and effortlessly and the human brain always tries to save energy. We use emotional response to somebody's look to assess how much we can trust them. That's not an optimal assessment for informed decision making. It certainly gives me to think that my opinion of political candidates probably depends on the shape of their nose. I really should go and read all these programs, comments and opinion pieces. But I have an essay to finish before the deadline, then write this overdue report and hurry to pick up the kids from daycare. Maybe I'll look up these candidates next week. Or the week after that. If only information wasn't so costly. If only it wouldn't take up so much time and energy.

But now imagine you could look at a candidate and in fact get a simple, fast, sensorial or emotional feedback how certain selected priorities and interest of this person match with yours. This would dramatically lower the cost of information. It would bring close that what is far away.

The ingredients for closing this feedback loop already exist, they just aren't combined suitably. Above I mentioned gamification to bridge long time distances. Other applications that make information less costly in terms of time and energy are sites dedicated to help you decide which party to vote based on answers to a set of questions, or dating sites that match your interests with potential mates. It's the same mechanism, but too scattered and not broad enough. The more dispersed these applications are, the more effort it takes to use them and the more costly the information becomes. We need it all in one place.

Concretely the feedback loop would work like this:

1. A user creates a personal priority map. In the future this may be done by a brain scan or by analyzing information transmitted from neural implants. Presently, questionnaires and other records must stand in. The questionnaires would cover for example personal values, various aspects of health and social life, political attitudes and personal taste. This should also include users' tolerance for risk and uncertainty because this is relevant to assess how good a match will be. This priority map is personal data that the user can update and expand, and share or make public selected parts of it.
2. Institutions that gather knowledge about the system (statistics, trends, predictions) make it available to users as correlations between actions and individual priorities. In return they use the shared parts of users' maps to obtain better information about the system, notably tensions that arise when priorities conflict, which can indicate problems with the current organization of the system.
3. Whenever a user takes a decision whose impact is likely to exceed the natural human ability to foresee consequences of individual actions he or she consults the priority map. The user can then tell how well a decision matches their recorded priorities and take this into account without having to bother with the details in every single case. The decisions serve to adapt the system.

This consultation of the priority map thus remedies the lack of intuition humans have for the behavior of highly connected and dynamic social networks. The goal is that