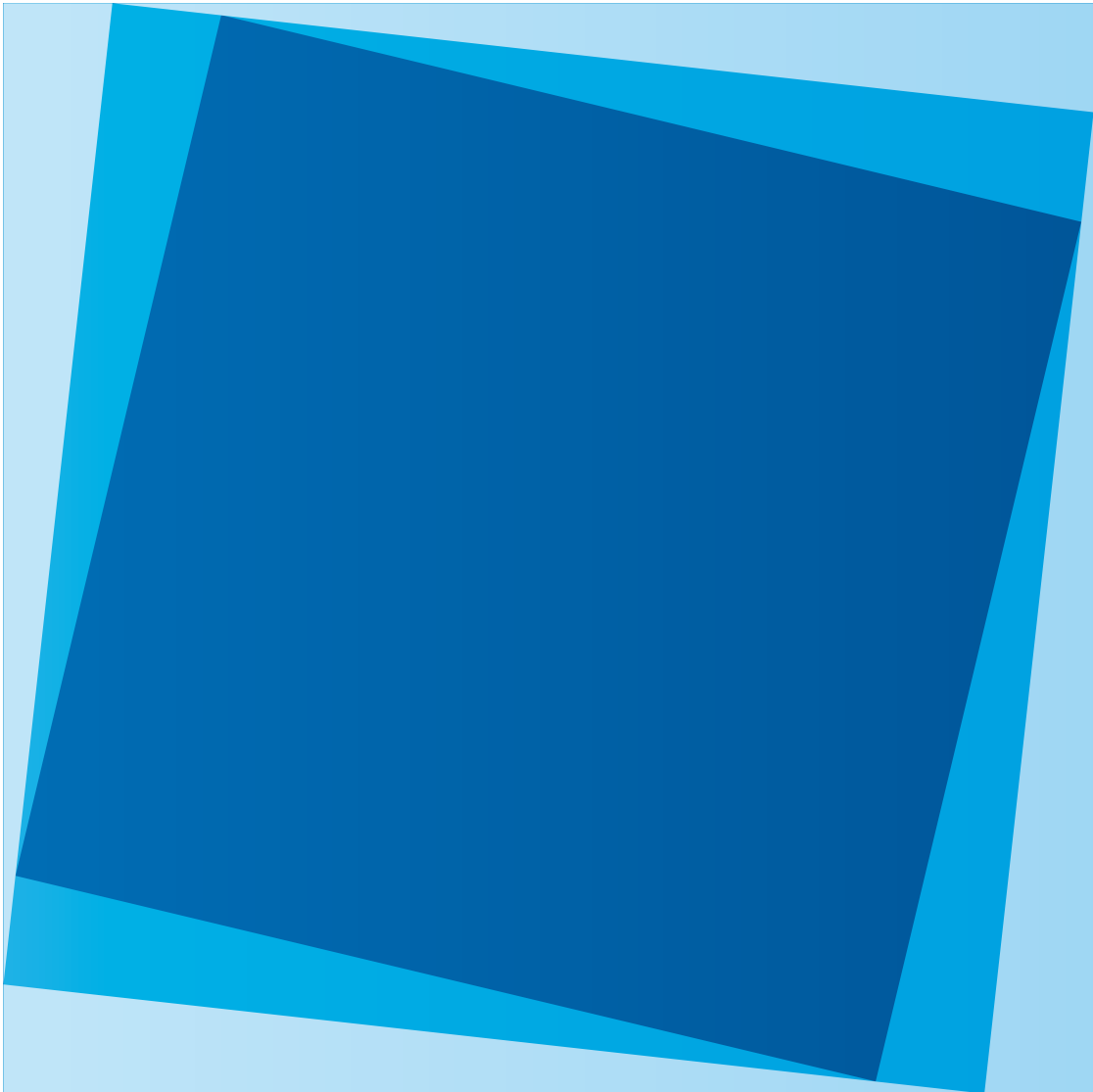


# **Towards a New Enlightenment?**



**A Transcendent  
Decade**

## Preface

This book, *Towards a New Enlightenment? A Transcendent Decade*, is the eleventh in an annual series that BBVA's OpenMind project dedicates to disseminating greater knowledge on the key questions of our time.

We began this series in 2008, with the first book, *Frontiers of Knowledge*, that celebrated the launching of the prizes of the same name awarded annually by the BBVA Foundation. Since then, these awards have achieved worldwide renown. In that first book, over twenty major scientists and experts used language accessible to the general public to rigorously review the most relevant recent advances and perspectives in the different scientific and artistic fields recognized by those prizes.

Since then, we have published a new book each year, always following the same model: collections of articles by key figures in their respective fields that address different aspects or perspectives on the fundamental questions that affect our lives and determine our future: from globalization to the impact of exponential technologies, and on to include today's major ethical problems, the evolution of business in the digital era, and the future of Europe.

The excellent reaction to the first books in this series led us, in 2011, to create OpenMind ([www.bbvaopenmind.com](http://www.bbvaopenmind.com)), an online community for debate and the dissemination of knowledge. Since then, OpenMind has thrived, and today it addresses a broad spectrum of scientific, technological, social, and humanistic subjects in different formats, including our books, as well as articles, posts, reportage, infographics, videos, and podcasts, with a growing focus on audiovisual materials. Moreover, all of the content is presented in Spanish and English, absolutely free of charge.

One of OpenMind's fundamental assets is its group of three hundred authors and collaborators: world-class specialists in their respective fields. Another is its community; in 2018, OpenMind will be visited by some seven million times by five-and-a-half million users from all over the world and two-hundred thousand followers on social media. Their participation, comments, and reposting of our contents bring life to the community.

In recent years, we have dedicated very particular interest to the technological revolution and its profound impact on all aspects of our lives. In our 2017 book, *The Next Step: Exponential Life*, we analyzed how this revolution is producing enormous changes in the economy, politics, society, culture, values, and everyday life. All this affects our understanding of what even humanity is, as technologies emerge that are capable of enormously enhancing human's physical and mental capacity and lifespan. Even our position as the planet's only intelligent species is brought into question by our coexistence and possible merging with increasingly intelligent machines.

All of this ushers in a new phase that led us to title last year's book *The Age of Perplexity*. Perplexity in the face of changes for which we have no guides or criteria for behaving—changes that call into question the very bases of our economic and political system. In the present book, *Towards a New Enlightenment? A Transcendent Decade*, we take another step forward, reviewing the most important changes that have occurred over the last ten years (which correspond to our project's current duration) and, on the basis of their analysis, look to the future in order to understand where they are leading and what decisions we must make on both individual and collective levels.

To approach this complex task, we have drawn on a group of twenty-three globally renowned, top-class specialists, and I wish to thank them all for their excellent collaboration and, in general, their generous support for our OpenMind project.

This book begins with a review of the recent advances and future lines of scientific development, as well as their technological applications. These advances are determining our future possibilities in cosmology and physics, anthropology and data science, nanotechnology, artificial intelligence and robotics, biotechnology, the social sciences, and so on.

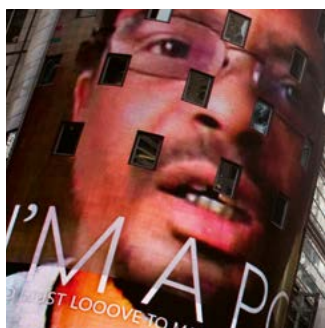
This knowledge helps us to better understand the trends shaping the principal concerns of our time: climate change and globalization, the economy and development, political organization and information, gender, the problems of massive urbanization, and cybersecurity, all of which are subsequently discussed.

From their analysis we can draw two conclusions: first, that the technological revolution's ultimate impact on human life, including the preservation of the environment, fundamentally depends on the decisions we begin making today; and second, that a key decision will be to foster what this book calls a "New Enlightenment," a broad dialogue to establish new philosophical and ethical bases for an economy, society, culture, and regulations adapted to our new scientific and technological surroundings, which will maximize their positive effects on growth and well-being, promote better distribution, and favor the development of shared initiatives to deal with global warming, environmental deterioration, and biodiversity. This will be a long and difficult task, but the sooner and more decisively we begin, the more likely we are to succeed. The alternative—failure—is unthinkable, because it opens the door to dystopias that some are already auguring and many already fear.

**Francisco González**

Group Executive Chairman, BBVA

Francisco González 7—25  
**Towards a New  
Digital Enlightenment:  
The Financial  
Industry's Role**



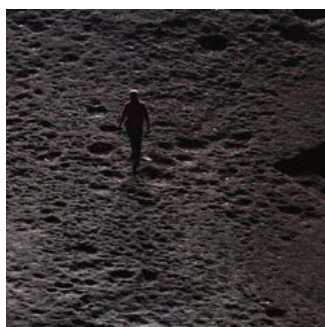
Martin Rees 26—44  
**The Past Decade  
and the Future  
of Cosmology and  
Astrophysics**



José Manuel Sánchez Ron 45—71  
**When Past Becomes  
Future: Physics in the  
21st Century**



María Martínón-Torres 72—84  
**Anthropology: What  
We Have Learned  
over the Last Decade**



Alex Pentland 85—105  
**Data for a New  
Enlightenment**



Sandip Tiwari 106—126  
**The Ghost in the  
Machine? Nanotechnology,  
Complexity, and Us**





Joanna J. Bryson

127 — 159

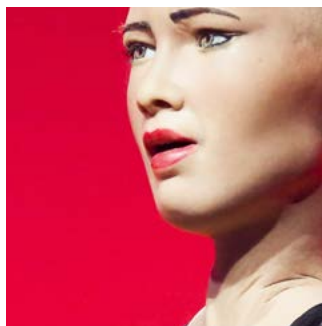
**The Past Decade  
and Future of AI's  
Impact on Society**



Ramón López de Mántaras

160 — 174

**The Future of AI:  
Toward Truly Intelligent  
Artificial Intelligences**



José M. Mato

175 — 187

**Turning Knowledge  
into Health**



Daniela Rus

188 — 202

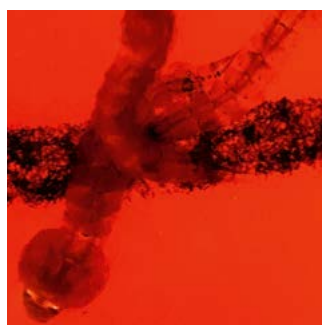
**A Decade of  
Transformation  
in Robotics**



Samuel H. Sternberg

203 — 219

**The Biological  
Breakthrough of  
CRISPR-Based  
Gene Editing**



Peter Kalmus

220 — 246

**Climate Change:  
Humanity at  
a Crossroads**



Ernesto Zedillo Ponce de León

247 — 265

**The Past Decade  
and the Future of  
Globalization**



Victoria Robinson

266 — 278

**Gender Inequalities:  
“Past” Issues and  
Future Possibilities**



Barry Eichengreen

279 — 295

**The Past Decade  
and the Future of  
the Global Economy**



Michelle Baddeley

296 — 310

**Behavioral Economics:  
Past, Present, and  
Future**



Nancy H. Chau and Ravi Kanbur

311 — 325

**The Past, Present,  
and Future of Economic  
Development**



Vivien A. Schmidt

326 — 346

**The Past Decade and  
the Future of Governance and  
Democracy: Populist Challenges  
to Liberal Democracy**



Diana Owen

347—365

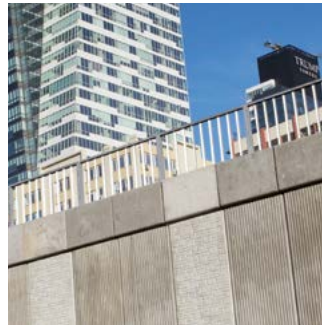
**The Past Decade  
and Future of Political  
Media: The Ascendancy  
of Social Media**



Yang Xu and Carlo Ratti

366—380

**Conquer the  
Divided Cities**



Amos N. Guiora

381—407

**Cybersecurity:  
A Cooperation Model**





**Francisco González**  
Group Executive Chairman,  
BBVA

Francisco González graduated in Economics and Business Administration from the Complutense University in Madrid. He has been Executive Chairman of BBVA since 2000 and is a Member of the Board of Directors of the Institute of International Finance (IIF), where he is also a Member of its Executive Committee. He is a Member of the European Financial Services Round Table (EFR), the Institut International d'Études Bancaires (IIEB), the International Advisory Panel of the Monetary Authority of Singapore (MAS), and the International Business Council (IBC), of the World Economic Forum (WEF), among other international fora. He is also a member of the Board of Trustees and a member of the Global Advisory Council of The Conference Board (TCB). He represents BBVA in the International Monetary Conference (IMC). He is also Chairman of the BBVA Foundation. Prior to the merger between Banco Bilbao Vizcaya and Argentaria, Francisco González was Chairman of Argentaria from 1996 to 1999, where he led the integration, transformation, and privatization of a diverse group of state-owned banks. He began his professional career in 1964 as a programmer in an IT company. His ambition to transform twenty-first-century banking with the support of new technologies dates back to this time.

Recommended book: *The Age of Perplexity: Rethinking the World We Knew*, OpenMind/BBVA, 2017.

**The last decade has been a turbulent one. The financial crisis and uncertainty about the effects of globalization and the technological revolution have led to broad questioning of the global order of representative democracy and free markets. This article argues that to materialize the enormous potential growth and well-being for all offered by the technological revolution, and to dissipate the climate of discontent and pessimism, we need a New Enlightenment: a renewal of our philosophical and ethical bases, and of our political, economic, and legal architecture. This may be a long and complex process, but some steps can be taken immediately through economic and legal reforms. The digital transformation of the financial industry is one of the reforms capable of improving productivity and driving more inclusive growth.**



Ours are turbulent times, where political, economic, and social patterns that seemed sturdy and practically permanent are being radically questioned.

Undoubtedly, this is largely due to the financial crisis that has marked the last decade. Eighty years after the Great Depression, and following the greatest known period of global prosperity that began at the end of World War II and ended with the “Great Moderation” at the beginning of our century, the world has again experienced a profound and worldwide economic crisis.

This crisis, with its ensuing unemployment, rapid deterioration of public accounts, and austerity policies, has particularly affected the most developed countries. There, the drop in production and income was greater and more long lasting, and social-policy cutbacks had a more powerful impact on their inhabitants, who have traditionally been much more protected and caught up in a process of rapid aging. Doubts about the sustainability of the welfare state, which had already arisen earlier and had led to liberal-oriented reforms in many countries, have grown stronger.

Moreover, these doubts have spread to other key aspects of the liberal democratic system. This has led to a proliferation of populist political options and proposed authoritarian solutions that have weakened citizens’ participation and trust in institutions and in the practice of politics and democratic representation. Populist solutions unfettered by institutions or traditional party politics are on the rise as a supposed representation of the “people” or the “real people” (Müller, 2017).

The result is increased political polarization, with a much more biased and less transparent debate that focuses more on the very short term and less on problem solving; more on the fabrication of enemies and confrontation rather than the quest for agreement.

At the same time, political information and communications are deteriorating, social networks favor media fragmentation, and polarization adds “spin” to the news or to deliberate forms of disinformation such as “fake news.”

## **Ours are turbulent times, where political, economic, and social patterns that seemed sturdy and practically permanent are being radically questioned. Undoubtedly, this is largely due to the financial crisis that has marked the last decade**

In this setting, citizens tend to feel less safe and more pessimistic; they turn to more simple and drastic solutions, assuming much more defined and monolithic national, ethnic, and religious identities, among others.

The result is a less cohesive and more divided society that shows greater hostility toward those perceived as “different” or “other”—especially immigrants.

And yet, these doubts and fears (of the *other*, of the future) did not begin with the crisis, and, in fact, they have continued to grow, even with the recovery of global growth in recent years. In the 1990s, the great sociologist Zygmunt Bauman (1998) coined the term “Unsicherheit” for the combination of uncertainty, insecurity, and vulnerability that he perceived in contemporary developed societies and that he attributed to the economic, social, and cultural effects of globalization and its difficult assimilation in those contexts—national, regional



and local—closest to the people. Moreover, even before the crisis, economists such as Mary Kaldor (2004) explained phenomena such as the so-called “new nationalisms” as reactions to globalization.

Many of the new political and economic proposals certainly define themselves as explicitly anti-globalization. Globalization is being questioned, and that very questioning has reached the geopolitical framework in which that phenomenon has developed.

At the end of the Cold War, the United States emerged as the only superpower, the guardian of an increasingly open and interconnected (globalized) world in which the liberal democratic regime seemed to have prevailed and a world order of increasingly integrated democracies and market economies would evolve.

But this view of the future is increasingly less certain. For two main reasons, the United States’ hegemony is no longer so clear: first, because that nation’s weight in the global economy is waning as other areas, especially China, expand much more rapidly. And second, since Trump’s election, the United States has adopted a more unilateralist approach that focuses on its most immediate interests at the expense of its former role as champion of democracy and free trade.

At the same time, supranational initiatives aimed at cooperation and economic integration (including the most ambitious of all: the European Union), international accords and organisms (such as the WTO, which is fundamental for sustained globalization), and organizations for coordinating global policies (the G8 and the G20) are growing weaker.

But this is not only a matter of globalization. Its repercussions (attributed or affirmed, positive or negative) are increasingly intertwined with those of the ongoing technological revolution. Technological advances drive globalization through improvements in telecommunications and greater connectivity, and the global setting is home to the technological revolution. That is where it can develop and make the most of its potential. In fact, throughout history, periods of globalization have normally been associated with accelerated technological progress.

The current period of globalization began after World War II, following the Great Depression of 1929 and the protectionist and nationalist reactions sparked by it (and here, it is certainly possible to recognize a disturbing parallel to the crisis of 2008 and our situation today). The so-called information revolution began at almost the same time as the global political and economic reorientation, and its ceaseless acceleration is ongoing.

This has been a period of unprecedented global prosperity; the world’s population has tripled in number and living conditions have vastly improved on most of the planet, with a considerable reduction in the number of persons living in poverty (World Bank, 2018).

It is impossible not to attribute much of the world economy’s outstanding performance to globalization and the technological revolution, as well as to the strengthening of institutions in many emerging countries. The extension of free-market principles, the rule of law, and the improvement of legal protection have helped many countries, especially in Asia, to make unprecedented advances in development, thus leading and driving global growth. As has always occurred in the past, such periods of globalization and technological progress have increased prosperity and created more jobs than they have destroyed (Mokyr et al., 2015).

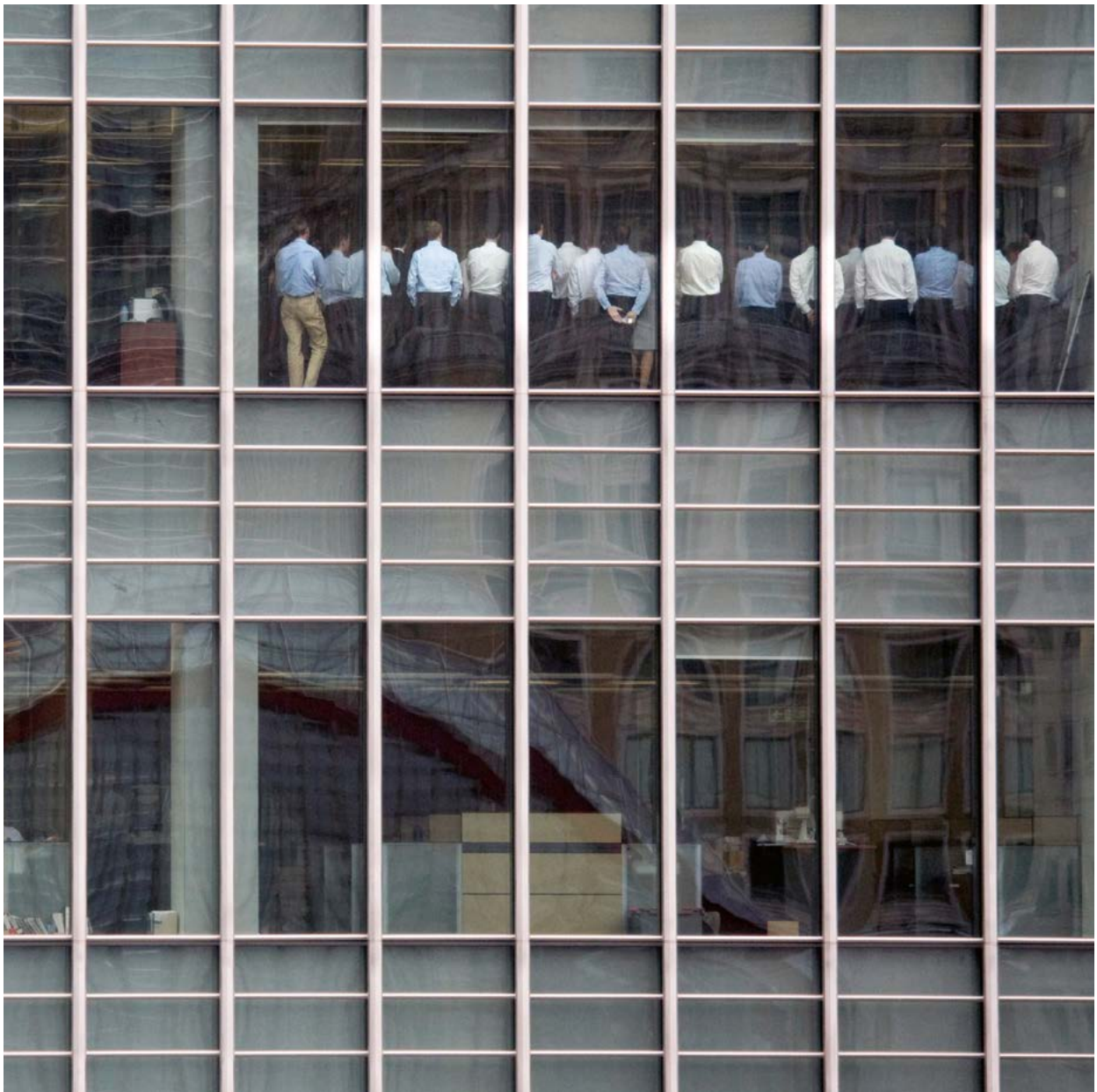
Why, then, is there such a sense of uncertainty, frustration, and pessimism?

As we mentioned above, the crisis and its aftereffects are part of the explanation. But if they were the only cause, we would be experiencing temporary phenomena which should have improved as the post-crisis recovery took shape. It is, therefore, necessary to look deeper into the outcomes of globalization and technological progress, one of the clearest of which is increased migratory flows.





Emergency meeting at Lehman Brothers' London office in that city's Canary Wharf financial district on September 11, 2008, four days before the company's bankruptcy and the start of the worldwide financial crisis





Emigration to developed countries is necessary, and even essential for sustaining growth and social-welfare systems. And yet, immigrants are frequently viewed as “unfair” competition for local employment. They are accused of contributing to the maintenance of low salaries and of a disproportionate use of social services.

At the same time, while globalization and technological advances have had no negative effects on income or aggregate employment in developed countries, they *have* affected its makeup and distribution.

Employment and salaries have particularly dropped in the manufacturing sector, first because many jobs have been outsourced to emerging countries where salaries are lower; and second, because automation and digitalization have made many of that sector’s routine and repetitive jobs redundant.

There has, however, been an increase in employment in the service sector, which is more difficult to automate. Most of this work, however, is unskilled, poorly paid, and open to competition from immigrants (Autor and Salomons, 2017). Simultaneously, the instability of the job market and greater job turnover has generated a growing portion of part-time, temporary, or freelance jobs in what has come to be known as a “gig economy” (Katz and Krueger, 2016).

Weak salary growth in developed countries is one of the most clearly established effects of globalization and technological progress. On the other hand, the number (and most of all, the remuneration) of highly skilled jobs has grown. Moreover, strong increases in productivity, as well as economies of scale and the network economies of the most digitalized sectors that drive the emergence of global monopolies, have led to major accumulations of income and wealth by very reduced segments of the population.

## **Stagnating salaries and increased inequality in developed countries, caused by globalization and technological change, as well as concern about the future of jobs in light of what has happened in many sectors, are at the base of the current climate of uncertainty and pessimism**

In summary, the distribution of wealth generated by globalization and technological advances has been very unequal. The winners are the richest sector of the population in both developed and emerging countries, as well as workers and the new middle classes in many emerging countries—principally China and India. The losers are the poorest of the poor (fundamentally in areas such as Sub-Saharan Africa) and the working and middle classes in developed countries and in many countries from the former communist block (Milanovic, 2016).

Stagnating salaries and increased inequality in developed countries, caused by globalization and technological change, as well as concern about the future of jobs in light of what has happened in many sectors, are at the base of the current climate of uncertainty and pessimism (Qureshi, 2017).

### **A New Society for the Digital Era**

Discontent with what has already occurred is accompanied by the anxiety (perplexity) generated by the speed and magnitude of scientific and technological advances. The bioscience and digital revolutions emerge as forces capable of transforming not only our economy and society but even our bodies and minds.





The Industrial Revolution was based on machines capable of surpassing the physical limitations of humans and animals. Today's revolution employs digital and biotechnologies that surpass not only our physical but also our intellectual limitations, and even the natural limitations of our lifespans. Sooner, rather than later, all of this will impose a radical rethinking of our economy, society, and culture, as well as of our ethical principles and even the fundamental philosophical bases of our existence as individuals and as a species.

## **The bioscience and digital revolutions emerge as forces capable of transforming not only our economy and society, but even our minds and bodies**

It is certainly impossible to foresee the nature and depth of the changes that will result from the monumental and rapid technological revolution that has only just begun.

We can fear all sorts of dystopias, but we can also view the technological revolution as a great opportunity to improve the well-being of all of the world's citizens.

Throughout human history, economic progress and social welfare have gone hand in hand with technological advances, and there is no reason why the present occasion should be an exception.

Those positive effects, however, have only emerged after long and difficult periods of transition, with winners and losers. We may recall, for example, the misfortune of farmhands and small landowners, dismal wages and factory work, as well as the exploitation of children and of the colonies at the dawn of the first Industrial Revolution.

Moreover, all technological revolutions since the Neolithic era, which ushered in agriculture and villages, have required long and profound processes of change in different areas.

For example, the development of the first Industrial Revolution that began in England in the mid-eighteenth century required, at the very least, numerous earlier relevant technological innovations, not least of which was the printing press, three centuries earlier! Equally necessary were the great discoveries made between the late fifteenth century and the end of the sixteenth century, which increased the availability of resources to Western Europe and, most of all, changed its worldview. Another crucial element was the scientific revolution of the sixteenth and seventeenth centuries, when Bacon, Galileo, Newton, Leibniz, and many others fathered science as we know it today.

Political systems were changing at the same time, with the birth of the nation states and the transition to the earliest (and limited) parliamentary democracies, followed by the American and French Revolutions.

All of this shaped a radical departure from what had formerly been the dominant thinking. The new model, which adds up to what we now call the "Enlightenment," was a major step away from a basically religious and institutional mentality toward an approach based on reason and facts that recognized the individual rights of every person regardless of their status. This fundamental philosophical change sustained the processes of economic, political, social, cultural, judicial, and institutional modernization, among others and led, in turn, to the contemporary parliamentary democracies and free-market economies that, after decades of growing success and prevalence over the communist alternative, are now being questioned.

Today, we are witnessing a new scientific and technological revolution that has been called the Fourth Industrial Revolution (Schwab, 2016). We have the economic and human



Oxfam activists wearing masks with the faces of political leaders (in the photo: Vladimir Putin and former Italian prime minister Paolo Gentiloni) demonstrate at a G20 meeting in Hamburg, Germany, in July 2017





resources to insure its advancement, but in essence, it is operating on bases that correspond, essentially, to the industrial age.

These bases need to be updated to drive the Fourth Industrial Revolution, to channel it, limit its risks, maximize its benefits, and extend them to the global population as a whole. In short, we need a New Enlightenment to organize scientific and technological advances in a new philosophical framework, to help adapt our ethical criteria and to guide the necessary legal and political changes.

This is unquestionably a very complex process. It may take decades to complete, and it certainly surpasses the capacity of any state or supranational organization.

We are facing years of profound debate among different and often opposite views, but we must stimulate and accelerate this process as much as possible. How? By fostering dialogue and the confluence of hard science and technology with social science and the humanities, including philosophy, ethics, and even the arts. Today, we have the tools to make this debate truly transparent, global, and open to all who have something to contribute.

And while this process is underway, it is important to concentrate on what can already be revised and adapted to new situations: especially in two closely connected areas: the economy and the law.

### **A New Political Economy, a New Legal Architecture**

In the area of political economy, it is important, first, to encourage and boost the positive effects of digital technology with reforms that foster research, development, and innovation, support entrepreneurship, increase market transparency and competitiveness, and finally, favor the infrastructures needed for the deployment and adoption of digital technologies.

As we have seen above, the job market is another priority. It is necessary to develop better policies for addressing unemployment, with social protection that is adequate but does not discourage job seeking in an environment marked by high levels of turnover. Active policies are needed to favor workers' recycling and mobility. And it is also essential to modernize regulations to better address the questions posed by a much greater diversity of work situations, including part-time jobs, freelancers, and so on.

The most important element of all is certainly education, because that is the most powerful means of insuring equal opportunities and social mobility. We definitely need more and better education to close the currently growing inequality gap.

The technological revolution will undoubtedly require improved technical training, as well as the provision of complementary, rather than alternative, skills required by technological advances. And, of course, we must also push for continuous training and recycling.

But that is not all. It is not even the most important. We live and will continue to live in a world of accelerating changes. It is, therefore, fundamental for education to promote certain values and attitudes: courage in the face of change, entrepreneurial spirit, resilience, the capacity to adapt, and teamwork, among others.

The extraordinary advance of biotechnologies and information technologies also poses very complex challenges to existing regulatory and legal frameworks.

Facing these challenges requires finding a fundamental balance that makes it possible to control the risks associated with technology without unduly hindering innovation or limiting its positive effects. This should lead to improvements in productivity, growth, and quality of life, and should be coordinated, as much as possible, at a global level.

Potentially, the regulation of all these activities will need to be revised in all spheres of life; however, five closely related areas stand out as particularly pressing.



The first is privacy, and there are already important initiatives in this area, including the General Data Protection Regulation (GDPR). This is a fine first step which will have to be further developed and improved in the coming years. It only affects Europeans, however, and ultimately we need to advance on a global scale toward more precise definitions of individual rights with regard to personal data, and more straightforward and efficient mechanisms for protecting and enforcing those rights.

The second aspect is market power. Digital technologies involve enormous economies of scale and scope, with the consequent “natural” tendency toward monopolies. Today, for example, Apple and Google have a duopoly on the smartphone-operating systems market, while Facebook and Google dominate the digital publicity market and Amazon is becoming increasingly dominant in online distribution and data-center infrastructure. These are just a few examples of similar phenomena visible in many other sectors, including urban passenger transit, the distribution of audiovisual content, and so on.

In this respect, there is cause for concern. Are new technologies undermining the competitive structures that drove growth in the twentieth century? And beyond the phenomenon of the previously mentioned platforms, a variety of macroeconomic evidence points to a growing polarization of productivity and sales in different industries (Van Reenen, 2018). Production appears to be concentrated in a small number of companies with high profit margins.

It has also been argued, however, that new technologies are not so much weakening competitiveness as changing its mechanisms. In that sense, concentration would not be reducing incentives and opportunities for equal competition among all to become new, globally successful mega-companies. But how could that happen? The argument is that new technologies are easily scalable and replicable. That should favor the spread of innovations generated by any company (even small ones) and limit the abuse of power by Internet giants. In fact, while there have been some instances of abuse of market share, there has yet to be evidence of systematic collusion: the Internet giants compete with each other to capture clients and expand into new activities and markets.

The matter has yet to be resolved, but, in any case, the risk of a permanent drop in the efficiency of multiple markets should lead us to develop adequate policies for fostering competition in digital settings.

Another important aspect is the concentration of income and wealth in a very small group of people linked to large technology companies while overall salaries have grown very moderately, or not at all, for some years. This produces aggravation that grows when those large companies are seen to be minimizing their worldwide fiscal contribution by paying taxes where it is most convenient for them.

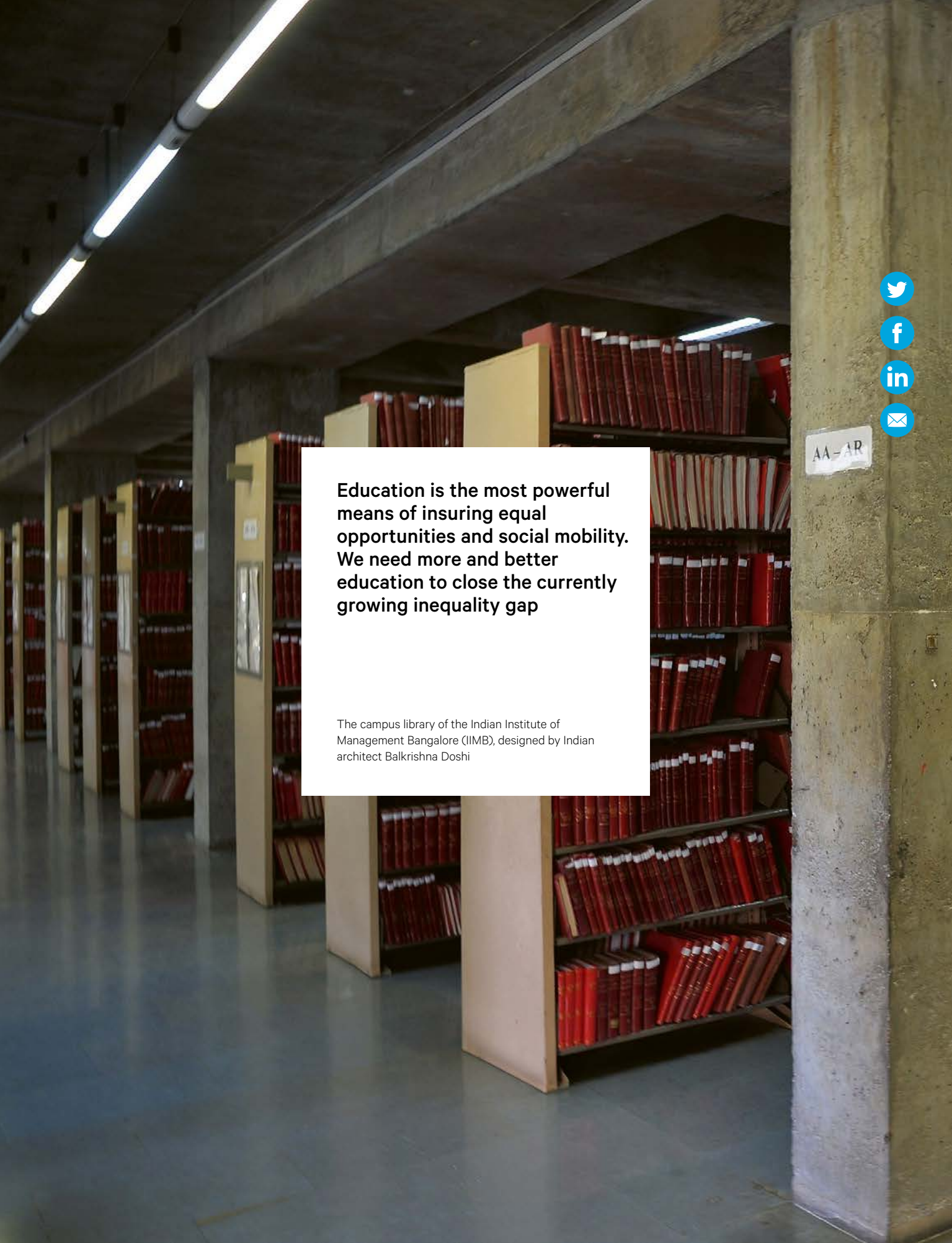
Control of these companies’ income and profits must certainly be improved, as must be the distribution of taxes among the different jurisdictions where they have clients and activity. Nonetheless, that requires a high degree of international coordination and the maintenance of adequate incentives to innovation.

It is also necessary to address the problems arising in the area of information. The emergence of new digital media, including social networks, has had a distorting effect on the transmission of news and on its transparency.

Finally, high priority must be assigned to cybersecurity, from crimes committed through social networks to threats to countries’ national security.

All of these elements will have to be taken into account in future technology policies and their regulation. This is an extraordinarily complex task, but it is fundamental if we want the technological revolution to generate prosperity and well-being commensurate with its potential.





**Education is the most powerful means of insuring equal opportunities and social mobility. We need more and better education to close the currently growing inequality gap**

The campus library of the Indian Institute of Management Bangalore (IIMB), designed by Indian architect Balkrishna Doshi











At a municipal vote in Seattle, various demonstrators hold signs demanding that locally based multinational corporations, such as Amazon, be taxed to combat rising housing prices





Beyond regulation, which will inevitably lag behind technological and business developments, we need to define and extend an ethical approach to the generation and application of scientific and technological advances. That approach needs to include moral and cultural values in the development of biotechnological applications and artificial intelligence.

If, for example, we believe that the democratic political system deserves to last through coming generations, then the information systems used by democratic governments must be designed to favor human rights, pluralism, separation of powers, transparency, fairness, and justice. Likewise, we must be certain that artificial intelligence algorithms used for contact with people in any business or activity are unbiased, do not discriminate against certain groups and do not unduly interfere in their freedom of choice.

For all these reasons, we need to change the manner in which we work with technology, generating more open, participative, and multidisciplinary ecosystems where more people can contribute ideas, talent, and resources.

We undoubtedly have the capacity to develop technology that does not enslave us and instead helps us to live better. But for this to happen we need to build ethical values into the design of that technology.

If we manage to do so, the Fourth Industrial Revolution may effectively become much less disturbing and more inclusive than the earlier three, a source of prosperity and well-being for all.

### **New Banking for a New Society**

Reaching this goal would be considerably easier with a “digital” financial system. Its far greater agility and efficacy would help improve people’s lives and contribute to much more inclusive growth.

Digital technologies have huge potential for driving this transformation, with enormous benefits for individual consumers and companies in terms of product quality, variety, convenience, and price. They will also allow thousands of millions of people from the lowest layers of society around the world to obtain access to financial services, thus increasing their possibility to prosper.

The last decade has been very difficult for banking. Following the extremely harsh impact of the financial crisis, including the failure of many banks and very negative effects on that industry’s reputation, banking now faces a period of reduced business growth and lower profitability in an environment characterized by low interest rates and greater capital requirements.

Clients, too, have changed. They demand different products and services and new ways of accessing them. And thousands of new service providers (start-ups or, for some services, major digital companies) are already meeting these demands (Lipton, Shrier, and Pentland, 2016).

Therefore, banking’s digital transformation is a case of both convenience for the common good and the needs of the sector itself.

The panorama in which this transformation will occur is very complex. On a global level, banking is growing more fragmented due to the annual entry of hundreds of competitors in a field that already includes 20,000 banks, worldwide.

At the same time, the banking industry is splintering. The immense majority of new competitors are breaking banking’s value chain, offering very specialized products and services.

But this trend will have to undergo an about-face in the future. First, because the banking sector already suffered from overcapacity in the past, and this is now growing even more acute. It is, therefore, likely that many banks will disappear, along with a multitude of start-ups whose mortality rate is always very high.





On the other hand, user-convenience calls for complete and integrated solutions, and that, in turn, points to supply side regrouping.

In light of what has occurred in other sectors, this regrouping will very likely take the form of platforms where different service providers compete—and they may frequently cooperate as well—to improve their offers to clients (González, 2017). The “owners” managing these platforms will control information generated by transactions and access to the end users. This control represents a major source of value.

What sort of companies will attain this position? Possibly some especially successful start-ups, certainly some of today’s major digital companies, and probably a few banks capable of transforming and adapting to this new scenario.

Banks have a good opportunity to successfully compete because they know the business, they are accustomed to operating in regulated environments, and have client confidence in the delicate area of handling their money. On the basis of that confidence and the client information they possess, they can build a platform that includes many more services. But the necessary transformation will not be possible for all banks, and the competition will be severe.

That competition is precisely what can lead to a much better financial system—one that is more efficient and productive, capable of providing better solutions to a larger number of users, including the hundreds of millions who currently have no access to financial services, and therefore capable of supporting growth and increased well-being for all.

## **On the basis of the confidence and the client information they possess, banks can build a platform that includes many more services. But the necessary transformation will not be possible for all**

Of course, in the financial world, as in others, materializing technology’s positive impact depends largely on the decisions we make, both as private agents and as public powers (especially, in the latter instance, regulators and supervisors).

Clearly, while technological advances offer great opportunities, they also involve risks. Grave disturbances of the financial system have a very important negative impact on growth, employment, and well-being, which is precisely why the financial system has been subject to particularly strict and detailed regulation.

On this occasion, financial regulators face an especially difficult task. First, because *digital* means *global*, and the new regulatory framework requires a far greater degree of international uniformity than presently exists. Second, because in a setting characterized by downturns in prices and margins of the sort imposed by a digital world, the consolidation and regrouping processes mentioned above, new competitors unaccustomed to operating in highly regulated environments, continually evolving business models and relentless technological change, the number of businesses that fail will greatly increase, as will the probability of systemic problems (Corbae and Levine, 2018).

The present system of financial regulation and supervision, which focuses on consumer protection and capital and liquidity requirements, is not adequate for handling the challenge of digital innovation faced by this industry in three main areas: consumer protection, financial stability, and the maintenance of a balanced framework for competitiveness.

New financial regulation must be constructed on very different bases, and rather than focusing on certain types of institutions, it must focus on activities and the risks they entail:



a wholistic regulation that also considers data protection, cybersecurity, and competition. To do so, it must also approach matters from all relevant angles (technological, legal, financial, and competitive).

From the very start, such transversal regulation based on close cooperation among authorities from different sectors and countries must also involve the private sector in its design process. And finally, it must be capable of adapting to an environment in constant evolution, both technologically and in terms of business models.

This is certainly a highly ambitious model but it can be approached in a pragmatic manner beginning with six priorities: data protection and access, the development of cloud computing for the financial industry, cybercrime, new developments in the area of payment, fostering the development of innovations in controlled environments (“sandboxes”), and finally, constructing a level playing field for both newcomers and established entities.

With regard to data, however, financial regulation is an extension of much more general regulation of overall personal data, including such delicate aspects as medical records. Its development must therefore be tied to progress in more general settings.

Various authorities, including the IMF, the Basel Committee on Banking Supervision, and the European Banking Authority, are already analyzing these questions, but their range is limited and partial, and in the face of insufficient global coordination, national authorities are approaching these subjects from different standpoints.

It is essential to open international debate on key subjects in order to define a shared body of principles. This will be a long and complex process in which progress will be slow, frequently partial, and always subject to revision as technology and business models evolve. It is, however, imperative to the construction of a better global financial system.

## **Rather than focusing on certain types of institutions, new financial regulation must focus on activities and the risks they entail: a wholistic regulation that also considers data protection, cybersecurity, and competition**

Moreover, and even more than new (and better) regulation, we need solid principles and ethical values. In this digital world, reputation and client confidence is as valuable an asset (or more so) as the finest technology.

These values must be fully integrated into the culture of entities seeking success in the digital age. And here, despite past errors and problems with their reputation, banks may have an advantage over competitors from other sectors: banks have always known the importance of reputation, and they have learned a great deal over the last few, very difficult years.

### **BBVA, Driving the Transformation of the Financial Industry**

BBVA began its digital transformation, or “long digital journey,” over ten years ago, in 2007, and we did not start from zero. From the very beginning, we had a very clear understanding that technology was going to radically transform our industry. It was not simply a matter of improving (even drastically) the efficiency of banks, or endowing them with new “remote” distribution channels. Instead, a bank had to become a different kind of company, capable of



competing in a completely new ecosystem with clients unlike their predecessors and other sorts of competitors.

At that point, we began to turn our vision into reality, undergoing a profound transformation that rests on three pillars: principles, people, and innovation.

We began acquiring the finest technology and we applied it to our operations, that is, to our business. But we never lost sight of the fact that technology is a tool for serving people: our clients and, of course, BBVA's staff.

We rapidly grasped that carrying out this transformation requires the finest talent, and the application of that talent and technology as a means of offering the best solutions to our clients and earning their trust to insure long-lasting relationships with them. They must trust in our technical confidence, but also in our will and capacity to treat them honestly.

Having the best staff and an excellent reputation among one's clients are key elements in banking, and that is why we have always assigned maximum importance to the values of prudence, transparency, and integrity.

The years that have passed since we began this transformation have been very difficult ones for the banking industry. The crisis was followed by a period of reinforced regulation and supervision, with the corresponding demand for increased capital and a drop in the industry's profitability.

Our solid principles have allowed us to emerge from this difficult period with new strength and, unlike many of our peers, without having needed any injection of public capital—not even to increase capital due to the crisis.

At the same time, technological advances and their adoption by the banking industry have continued to accelerate. When we began our transformation, there were very few smartphones, and their capacities were vastly inferior to the present ones. Apple launched its first iPhone in 2007, and, in one decade, the telephone has become clients' leading means of interacting with their banks.

**At BBVA, we have been working on a profound cultural transformation. Our basic principles have not changed, but our working processes, organizational structures, and talents have certainly changed, as have the attitudes that needed to be promoted: a positive attitude toward change, flexibility, teamwork, and obsession with our clients and with the need to continually improve their experience**

This has also been the age of burgeoning cloud computing and big data, and, more recently, artificial intelligence and the distributed-ledger technologies underlying blockchains. These may become the bases for even more profound future transformations that are largely unforeseeable at the present time.

All of these developments have affected our project and required us to modify its original design (sometimes in highly significant ways).

Another key element is the configuration of our staff. Carrying out a pioneering project requires the finest talent, and for some time it was difficult to attract digital talent because conventional banking did not seem like an obvious destination. That, in turn, slowed progress and created a vicious circle. But with considerable effort we became a much more attractive





place to work, a successful mixture of digital and financial talent—the latter of which we already had in abundance. This has turned the situation around and produced a reciprocal situation in which the advances in our project made it more and more attractive to new talent capable of helping us to advance even farther. Of course, the finest staff requires excellent leadership, with directors capable of both sharing and driving the project.

Last, but not least, we have been working on a profound cultural transformation. Our basic principles have not changed, but our working processes, organizational structures, and talents have certainly changed, as have the attitudes that needed to be encouraged: a positive attitude toward change, flexibility, teamwork, obsession with our clients, and with the need to continually improve their experience, a focus on getting things done, the ambition to constantly grow better, to adopt the loftiest goals and to pursue them tenaciously.

Over the last decade, we have made many mistakes, with technology and with certain personnel, but we have learned from our errors and we have continued working. Today, we lead the industry in technology, but, most of all, we have the staff, the talent, and the agile organization, leadership, and culture needed for advancing at an ever-faster rate. The results have been very significant.

Our mobile banking application in Spain has been rated best in the world in 2017 and 2018 by Forrester (with Garanti, our Turkish bank's application, rated second in the world).

In June 2018, forty-six percent of our clients were digital and thirty-eight percent were mobile. Those numbers will surpass fifty percent in 2019. Moreover, digital sales are approaching fifty percent of our total.

More importantly, our digital clients are even more satisfied than our conventional ones, and this is largely the reason why our overall client satisfaction index continues to rise, making us leaders in the majority of the countries in which we operate.

Today, BBVA is at the forefront of the global financial system. There is still much to be done: technology changes constantly, new ideas arise, as do new business models and new and ever-stronger competitors. But this competition is what allows us to live up to our vision of “bringing the age of opportunity to everyone” on a daily basis, while contributing to the improvement of the global financial system.



## Select Bibliography

- Autor, D., and Salomons, Anna. 2017. “Does productivity growth threaten employment?” Article written for the BCE Forum on Central Banking, Sintra, June 2017.
- Bauman, Zygmunt. 1998. *Globalization. The Human Consequences*. New York: Columbia University Press.
- Corbae, D. and Levine, Ross. 2018. “Competition, stability, and efficiency in financial markets.” Paper presented at the Jackson Hole Economic Policy Symposium, August 23–25, 2018.
- González, F. 2017. “From the age of perplexity to the era of opportunities: Finance for growth.” In *The Age of Perplexity: Rethinking the World We Knew*, Madrid: BBVA OpenMind.
- Kaldor, Mary. 2004. “Nationalism and globalization.” *Nations and Nationalism* 10.1–2: 161–177.
- Katz, Lawrence F., and Krueger, Alan B. 2016. “The rise and nature of alternative work. Arrangements in the United States 1995–2015.” National Bureau of Economic Research.
- Lipton, A., Shrier, David, and Pentland, Alex. 2016. “Digital banking manifesto: The end of banks?” Cambridge, MA: MIT.
- Milanovic, Branko. 2016. *Global Inequality. A New Approach for the Age of Globalization*. Cambridge, MA: Harvard University Press.
- Mokyr, Joel, Vickers, Chris, and Ziebarth, Nicolas L. 2015. “The history of technological anxiety and the future of economic growth: Is this time different?” *Journal of Economic Perspectives* 29(3): 31–50.
- Müller, J. W. 2017. “The rise and rise of populism?” In *The Age of Perplexity: Rethinking the World We Knew*, Madrid: BBVA OpenMind.
- Qureshi, Zia. 2017. “Advanced tech, but growth slow and unequal: Paradoxes and policies.” In *The Age of Perplexity: Rethinking the World We Knew*, Madrid: BBVA OpenMind.
- Schwab, Klaus. 2016. “The fourth industrial revolution. What it means, how to respond.” World Economic Forum.
- Van Reenen, John. 2018. “Increasing differences between firms: Market power and the macro-economy.” Paper presented at the Jackson Hole Economic Policy Symposium, August 23–25, 2018.
- World Bank. 2018. Consulted October 6, 2018 at <http://databank.worldbank.org/data/reports.aspx?source=poverty-and-equity-database>.



**Martin Rees**  
University of Cambridge

Martin Rees is a cosmologist and space scientist. After studying at Cambridge University, he held various posts in the UK and elsewhere, before returning to Cambridge, where he has been a professor, Head of the Institute of Astronomy, and Master of Trinity College. He has contributed to our understanding of galaxy formation, black holes, high-energy phenomena in the cosmos, and the concept of the multiverse. He has received substantial international recognition for his research. He has been much involved in science-related policy, being a member of the UK's House of Lords and (during 2005–10) President of the Royal Society, the independent scientific academy of the UK and the Commonwealth. Apart from his research publications, he writes and lectures widely for general audiences, and is the author of eight books, the most recent being *On the Future* (2018).

Recommended books: *Universe*, Martin Rees, Dorling Kindersley, 2012; *On the Future*, Martin Rees, Princeton University Press, 2018.

In the last decade, there has been dramatic progress in exploring the cosmos. Highlights include close-up studies of the planets and moons of our Solar System; and (even more dramatic) the realization that most stars are orbited by planets, and that there may be millions of Earth-like planets in our Galaxy. On a still larger scale, we have achieved a better understanding of how galaxies have developed, over 13.8 billion years of cosmic history, from primordial fluctuations. These fluctuations might have been generated via quantum effects when our entire cosmos was of microscopic size. Einstein's theory received further confirmation with the detection of gravitational wave—a tremendous technological achievement. Future advances will depend on more powerful instruments, which could reveal evidence of life on exoplanets, and yield a better understanding of the big bang, and the ultimate fact of our cosmos.



Astronomy is the grandest of the environmental sciences, and the most universal—indeed, the starry sky is the one feature of our environment that has been shared, and wondered at, by all cultures throughout human history. Today, it is an enterprise that involves a huge range of disciplines: mathematics, physics, and engineering, of course; but others too.

Astronomers aim to map and survey all the various entities—planets, stars, galaxies, black holes, and so forth—that pervade the cosmos. We then want to use our knowledge of physics to understand the exotic objects that our telescopes have revealed. A more ambitious aim is to understand how the entire cosmic panorama, of which we are a part, emerged from our universe's hot, dense beginning.

The pace of advance has crescendoed rather than slackened; instrumentation and computer power have improved hugely and rapidly. The last decade, in particular, has witnessed some astonishing advances. And the promise for the future is even brighter—astronomy offers splendid opportunities for young researchers who want to enter a vibrant field.

In this paper I focus on three topics: firstly, planets and exoplanets, relatively “local” on a cosmic scale; secondly, gravity and black holes, the extragalactic realm; and then thirdly, and more speculatively, some concepts that aim to understand the cosmos as a whole.

## Planets, Exoplanets, and Life

Human spaceflight has somewhat languished in the decades since those of us who are now middle-aged were inspired by the Apollo program and the Moon landings. But space technology has burgeoned—for communication, environmental monitoring, satnav, and so forth. We depend on it every day. And for astronomers it has opened new “windows”: telescopes in space reveal the far infrared, the UV, X-ray, and gamma ray sky. Even though humans have not ventured further than the Moon, unmanned probes to other planets have beamed back pictures of varied and distinctive worlds.

Among highlights of the last decade, ESA's “Rosetta” mission landed a small probe on a comet—to check, for instance, if isotopic ratios in the cometary ice are the same as in the Earth's water. This is crucial for deciding where that water came from. NASA's “New Horizons” probe has passed Pluto, and is now heading into the Kuiper Belt, replete with minor planets.

**Among highlights of the last decade, ESA's “Rosetta” mission landed a small probe on a comet—to check, for instance, if isotopic ratios in the cometary ice are the same as in the Earth's water**

Rosetta took about a decade to reach its destination, preceded by almost that long in planning and construction. Its robotic technology dates from the 1990s—that is plainly frustrating for the team that developed the project, because present-day designs would have far greater capabilities. And the same is true for “New Horizons”—which nonetheless transmitted back to us high-definition pictures of Pluto, ten thousand times further from Earth than the Moon is. And the “Cassini” probe, which spent thirteen years exploring Saturn and its moons, is



even more of an antique: twenty years elapsed between its launch and its final plunge into Saturn in late 2017.

We are aware how mobile phones have changed in the last fifteen to twenty years—so imagine how much more sophisticated today’s follow-ups to these missions could be. During this century, the entire Solar System—planets, moons, and asteroids—will be explored and mapped by flotillas of tiny robotic craft, interacting with each other like a flock of birds. Giant robotic fabricators will be able to construct, in space, huge solar-energy collectors and other artifacts. The Hubble Telescope’s successors, with huge gossamer-thin mirrors assembled under zero gravity, will further expand our vision of stars, galaxies, and the wider cosmos. The next step would be space mining and fabrication. (And fabrication in space will be a better use of materials mined from asteroids than bringing them back to Earth.)

It is robots, and not humans, that will build giant structures in space. And sophisticated robots will explore the outer planets: they will have to utilize the techniques of deep learning and artificial intelligence (AI) to make autonomous decisions—the travel time for a radio signal to the outer planets is measured in hours or even days, so there is no possibility of direct control from Earth. These robots will not be humanoid in size and shape. Humans are adapted to the Earth’s environment. Something more spider-like would be better suited to the weaker gravity of Pluto or the asteroids.

But will these endeavors still leave a role for humans? There is no denying that NASA’s “Curiosity,” a vehicle the size of a small car that has, since 2011, been trundling across Martian craters, may miss startling discoveries that no human geologist could overlook. But machine learning is advancing fast, as is sensor technology; whereas the cost gap between manned and unmanned missions remains huge.

Robotic advances will surely erode the practical case for human spaceflight. Nonetheless, I hope people will follow the robots, though it will be as risk-seeking adventurers rather than for practical goals. The most promising developments are spearheaded by private companies. For instance, SpaceX, led by Elon Musk, who also makes Tesla electric cars, has launched unmanned payloads and docked with the Space Station. He hopes soon to offer orbital flights to paying customers.

**It is robots, and not humans, that will build giant structures in space. And sophisticated robots will explore the outer planets: they will have to utilize the techniques of deep learning and artificial intelligence to make autonomous decisions**

Indeed, I think the future of manned spaceflight, even to Mars, lies with privately funded adventurers, prepared to participate in a cut-price program far riskier than any government would countenance when civilians were involved—perhaps even one-way trips. (The phrase “space tourism” should definitely be avoided. It lulls people into believing that such ventures are routine and low-risk; and if that is the perception, the inevitable accidents will be as traumatic as those of the US Space Shuttle were. Instead, these cut-price ventures must be “sold” as dangerous sports, or intrepid exploration.)

By 2100, groups of pioneers may have established bases independent from the Earth—on Mars, or maybe on asteroids. But do not ever expect mass emigration from Earth. Nowhere in our Solar System offers an environment even as clement as the Antarctic or the top of Everest.





Space does not offer an escape from Earth's problems. Dealing with climate change on Earth is a doddle compared with terraforming Mars.

What are the long-term hopes for space travel? The most crucial impediment today stems from the intrinsic inefficiency of chemical fuel, and the consequent requirement to carry a weight of fuel far exceeding that of the payload. Launchers will get cheaper when they can be designed to be more fully reusable. But so long as we are dependent on chemical fuels, interplanetary travel will remain a challenge. A space elevator would help. And nuclear power could be transformative. By allowing much higher in-course speeds, it would drastically cut the transit times to Mars or the asteroids (reducing not only astronauts' boredom, but their exposure to damaging radiation).

The question that astronomers are most often asked is: "Is there life out there already? What about the 'aliens' familiar from science fiction?" Prospects look bleak in our Solar System, though the discovery of even the most vestigial life-forms—on Mars, or in oceans under the ice of Europa (one of Jupiter's moons) or Enceladus (a moon of Saturn)—would be of crucial importance, especially if we could show that this life had an independent origin.

But prospects brighten if we widen our horizons to other stars—far beyond the scale of any probe we can now envisage. The hottest current topic in astronomy is the realization that many other stars—perhaps even most of them—are orbited by retinues of planets, like the Sun is. These planets are not directly seen but inferred by precise measurement of their parent star. There are two methods:

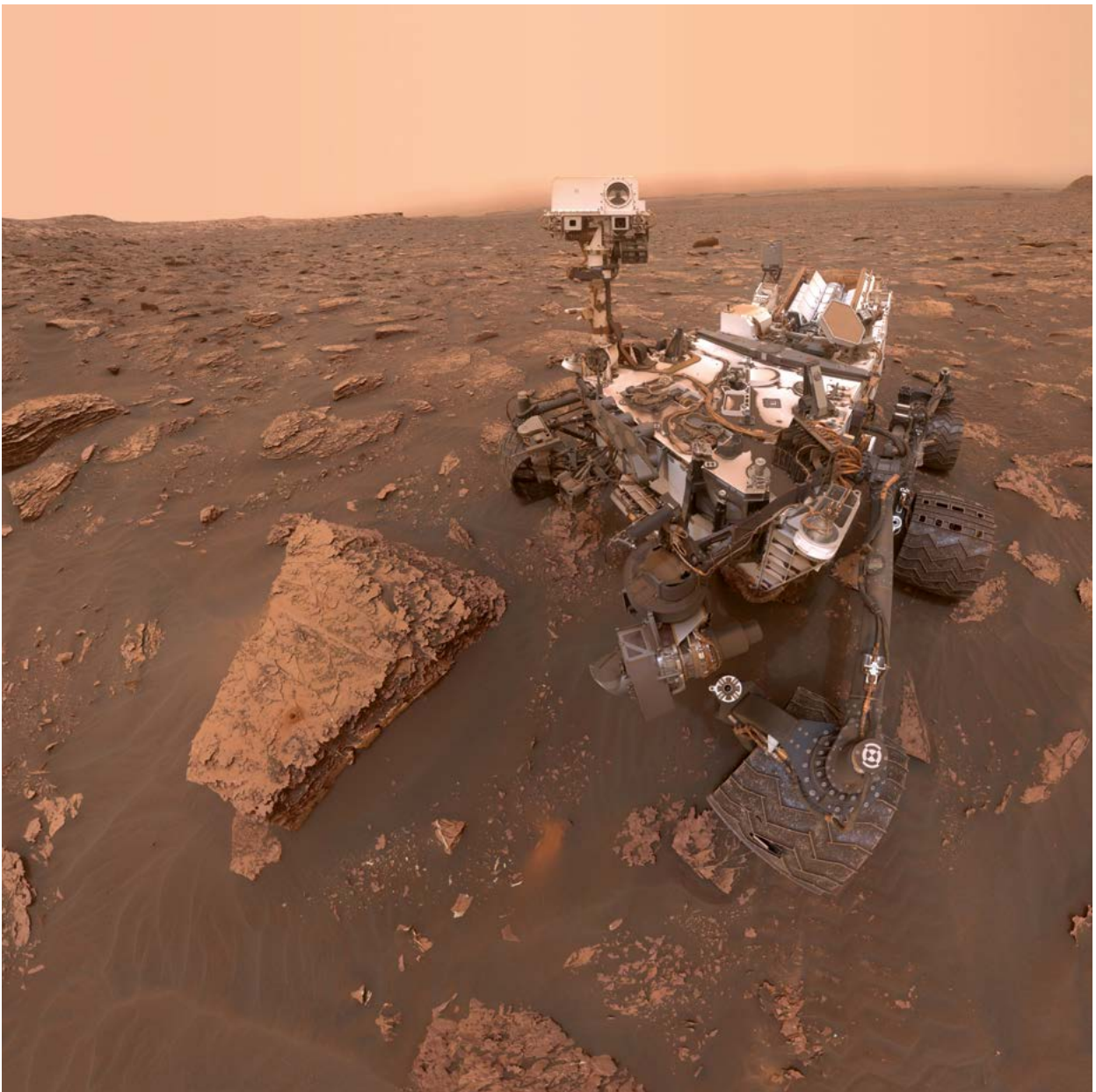
1. If a star is orbited by a planet, then both planet and star move around their center of mass—the barycenter. The star, being more massive, moves slower. But the tiny periodic changes in the star's Doppler effect can be detected by very precise spectroscopy. By now, more than five hundred exo-solar planets have been inferred in this way. We can infer their mass, the length of their "year," and the shape of their orbit. This evidence pertains mainly to "giant" planets—objects the size of Saturn or Jupiter. Detecting Earth-like planets—hundreds of times less massive—is a real challenge. They induce motions of merely centimeters per second in their parent star.

2. But there is a second technique that works better for smaller planets. A star would dim slightly when a planet was "in transit" in front of it. An Earth-like planet transiting a Sun-like star causes a fractional dimming, recurring once per orbit, of about one part in 10,000. The Kepler spacecraft was pointed steadily at a 7-degree-across area of sky for more than three years—monitoring the brightness of over 150,000 stars, at least twice every hour, with precision of one part in 100,000. It found more than two thousand planets, many no bigger than the Earth. And, of course, it only detected transits of those whose orbital plane is nearly aligned with our line of sight. We are specially interested in possible "twins" of our Earth—planets the same size as ours, on orbits with temperatures such that water neither boils nor stays frozen. Some of these have already been identified in the sample, suggesting that there are billions of Earth-like planets in the Galaxy.

The real challenge is to see these planets directly, rather than inferring them from observations of their parent star. But that is hard. To illustrate the challenge, suppose an alien astronomer with a powerful telescope was viewing the Earth from (say) thirty light-years away—the distance of a nearby star. Our planet would seem, in Carl Sagan's phrase, a "pale blue dot,"<sup>1</sup> very close to a star (our Sun) that outshines it by many billions: a firefly next to a searchlight. But if it could be detected, even just as a "dot," several features could be inferred. The shade



A self-portrait by NASA's Curiosity rover taken on Sol 2082 (June 15, 2018). A Martian dust storm has reduced sunlight and visibility at the rover's location in Gale Crater. A drill hole can be seen in the rock to the left of the rover at a target site called "Duluth"





of blue would be slightly different, depending on whether the Pacific Ocean or the Eurasian land mass was facing them. The alien astronomers could infer the length of our “day,” the seasons, the gross topography, and the climate. By analyzing the faint light, they could infer that it had a biosphere.

Within ten to fifteen years, the huge E-ELT (Europe’s “Extremely Large Telescope”), being built by the European Southern Observatory on a mountain in Chile—with a mosaic mirror thirty-nine meters across—will be drawing inferences like this about planets the size of our Earth, orbiting other Sun-like stars. But what most people want to know is: could there be life on them—even intelligent life? Here we are still in the realm of science fiction.

We know too little about how life began on Earth to lay confident odds. What triggered the transition from complex molecules to entities that can metabolize and reproduce? It might have involved a fluke so rare that it happened only once in the entire Galaxy. On the other hand, this crucial transition might have been almost inevitable given the “right” environment. We just do not know—nor do we know if the DNA/RNA chemistry of terrestrial life is the only possibility, or just one chemical basis among many options that could be realized elsewhere.

Moreover, even if simple life is widespread, we cannot assess the odds that it evolves into a complex biosphere. And, even it did, it might anyway be unrecognizably different. I will not hold my breath, but the SETI program is a worthwhile gamble—because success in the search would carry the momentous message that concepts of logic and physics are not limited to the hardware in human skulls.

## **The hottest current topic in astronomy is the realization that many other stars—perhaps even most of them—are orbited by retinues of planets, like the Sun is**

And, by the way, it is too anthropocentric to limit attention to Earth-like planets even though it is a prudent strategy to start with them. Science-fiction writers have other ideas—balloon-like creatures floating in the dense atmospheres of Jupiter-like planets, swarms of intelligent insects, and so on. Perhaps life can flourish even on a planet flung into the frozen darkness of interstellar space, whose main warmth comes from internal radioactivity (the process that heats the Earth’s core). We should also be mindful that seemingly artificial signals could come from superintelligent (though not necessarily conscious) computers, created by a race of alien beings that had already died out. Indeed, I think this is the most likely possibility.

We may learn this century whether biological evolution is unique to our Earth, or whether the entire cosmos teems with life—even with intelligence. We cannot give any firm estimates of how likely this is. Even if simple life is common, it is a separate question whether it is likely to evolve into anything we might recognize as intelligent or complex. It could happen often. On the other hand, it could be very rare. That would be depressing for the searchers. But it would allow us to be less cosmically modest: Earth, though tiny, could be the most complex and interesting entity in the entire Galaxy.

The E-ELT will reveal exoplanets, but still only as points of light. But by mid-century there may be huge mirrors in space that could actually resolve an image of an Earth-sized world orbiting another star. Perhaps some of these may have evidence of vegetation or other life. We have had, since 1968, the famous image—iconic ever since among environmentalists—of

our Earth, taken by Apollo astronauts orbiting the Moon. Perhaps by its centenary, in 2068, we will have an even more astonishing image: another Earth, even one with a biosphere.



### Strong Gravity and the Large-Scale Cosmos

Back now to the inanimate world of physics and chemistry—far simpler than biology. What has surprised people about the newly discovered planetary systems is their great variety. But the ubiquity of exoplanets was not surprising. We have learned that stars form via the contraction of clouds of dusty gas; and if the cloud has any angular momentum, it will rotate faster as it contracts, and spin off a dusty disk around the protostar. In such a disk, gas condenses in the cooler outer parts; closer in, less volatile dust agglomerates into rocks and planets. This should be a generic process in all protostars.

The crucial force that allows stars to form and holds planets in orbit around them is, of course, that of gravity. And it is Einstein's theory of general relativity that describes precisely how gravity behaves. Einstein did not "overthrow" Newton, but his theory applied more widely than Newton's and offered a deeper understanding of gravity in terms of space and time: in the words of the great physicist John Archibald Wheeler: "Spacetime tells matter how to move; matter tells spacetime how to curve."<sup>2</sup> This great theory was proposed in 1915. But for the first fifty years after its discovery relativity was somewhat isolated from the mainstream of physics and astronomy. The gravitational effects governing ordinary stars and galaxies were weak enough to be adequately described by Newtonian theory—general relativity was no more than a tiny correction.

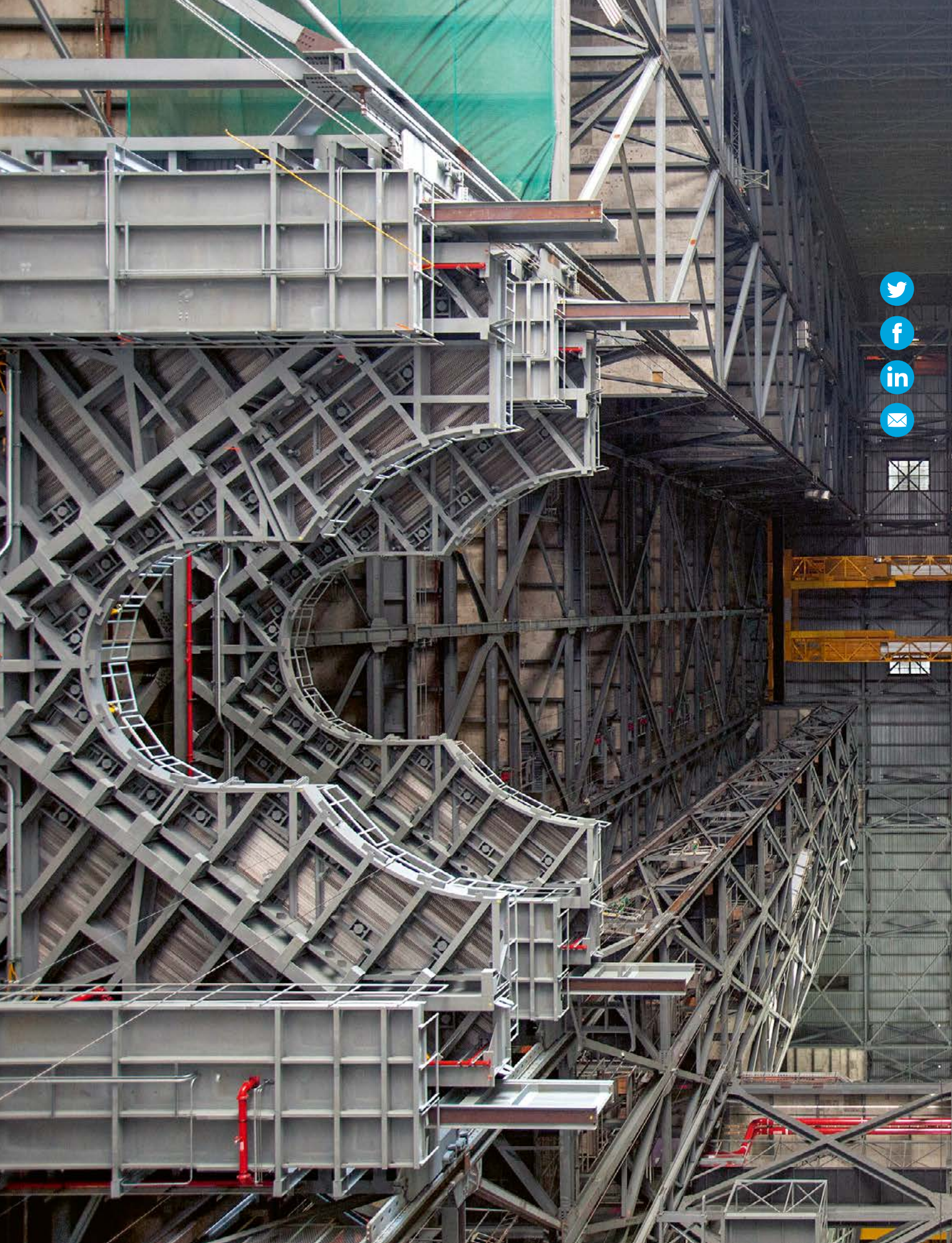
## What has surprised people about the newly discovered planetary systems is their great variety. But the ubiquity of exoplanets was not surprising

This situation changed in 1963 with the discovery of quasars—hyper-luminous beacons in the centers of some galaxies, compact enough to vary within hours or days, but which vastly outshine their host galaxy. Quasars revealed that galaxies contained something more than stars and gas. That "something" is a huge black hole lurking in their centers. Quasars are specially bright because, as we now recognize, they are energized by emission from magnetized gas swirling into a central black hole.

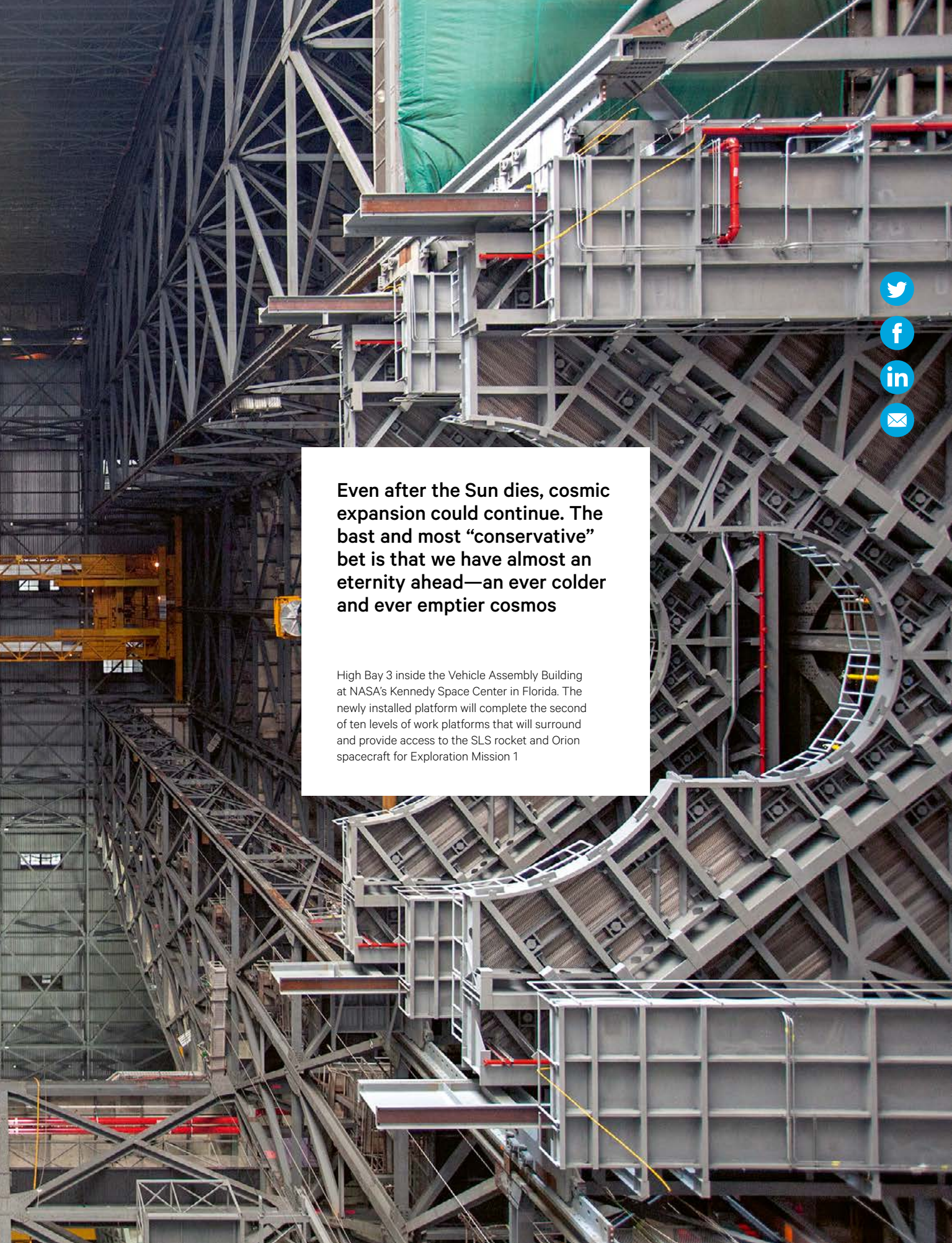
Quasars were a key stimulus to the emergence of "relativistic astrophysics." But not the only one. In particular, another surprise was the detection of neutron stars. One of the best-known objects in the sky is the Crab Nebula: the expanding debris from a supernova witnessed by Chinese astronomers in AD 1054. What kept it shining, so blue and bright, was a longtime puzzle. The answer came when it was discovered that the innocuous-seeming star in its center was anything but normal. It was actually a neutron star spinning at 30 revs per second and emitting a wind of fast electrons that generated the blue light. In neutron stars relativistic effects are ten to twenty percent—not merely a tiny correction to Newton.

Supernovae are crucial to us: if it was not for them we would not be here. By the end of a massive star's life, nuclear fusion has led to an onion skin structure—with hotter inner shells processed further up the periodic table. This material is then flung out in the supernova









**Even after the Sun dies, cosmic expansion could continue. The bast and most “conservative” bet is that we have almost an eternity ahead—an ever colder and ever emptier cosmos**

High Bay 3 inside the Vehicle Assembly Building at NASA's Kennedy Space Center in Florida. The newly installed platform will complete the second of ten levels of work platforms that will surround and provide access to the SLS rocket and Orion spacecraft for Exploration Mission 1



explosion. The debris then mixes into the interstellar medium and recondenses into new stars, orbited by planets.

The concept was developed primarily by Fred Hoyle and his associates. They analyzed the specific nuclear reactions involved, and were able to understand how most atoms of the periodic table came to exist and why oxygen and carbon (for instance) are common, whereas gold and uranium are rare. Some elements are forged in more exotic environments—for instance, gold is made in the cataclysmic collisions of neutron stars—a phenomenon not observed until 2017, when a gravitation wave signal, interpreted as a merger of two neutron stars, was followed up by telescopes that detected the event in many wavebands.



**Supernovae are crucial to us: if it was not for them we would not be here. By the end of a massive star's life, nuclear fusion has led to an onion skin structure—with hotter inner shells processed further up the periodic table. This material is then flung out in the supernova explosion. The debris then mixes into the interstellar medium and recondenses into new stars, orbited by planets**

Our Galaxy is a huge ecological system where gas is being recycled through successive generations of stars. Each of us contains atoms forged in dozens of different stars spread across the Milky Way, which lived and died more than 4.5 billion years ago, polluting the interstellar cloud in which the Solar System condensed.

The 1960s saw the first real advance in understanding black holes since Julius Robert Oppenheimer and his co-workers, in the late 1930s, clarified what happens when something falls into a black hole and cuts itself off from the external world. (And it is interesting to conjecture how much of the 1960s work Oppenheimer might have preempted if World War II had not broken out the very day—September 1, 1939—that his key paper appeared in the *Physical Review*.)

Theorists in the 1960s were surprised when their calculations showed that all black holes that had settled into a steady state were “standardized” objects, specified by just two numbers: their mass, and their spin—no other parameters. This realization hugely impressed the great theorist Subrahmanyan Chandrasekhar, who wrote that “in my entire scientific life, the most shattering experience has been the realization that an exact solution of Einstein’s equations ... provides the absolutely exact representation of untold numbers of massive black holes that populate the Universe.”<sup>3</sup>

A dead quasar—a quiescent massive black hole—lurks at the center of most galaxies. Moreover, there is a correlation between the mass of the hole and that of its host galaxy. The actual correlation is with the bulge (non-disk) component, not the whole galaxy. Our own Galaxy harbors a hole of around four million solar masses—modest compared to the holes in the centers of giant elliptical galaxies, which weigh billions of solar masses.

Einstein was catapulted to worldwide fame in 1919. On May 29 that year there was a solar eclipse. A group led by the Cambridge astronomer Arthur Eddington observed stars appearing close to the Sun during the eclipse. The measurements showed that these stars were displaced from their normal positions, the light from them being bent by the Sun’s gravity. This confirmed one of Einstein’s key predictions. When these results were reported at the



Royal Society in London, the world press spread the news. “Stars all askew in the heavens; Newton Overthrown” was the rather over-the-top headline in *The New York Times*.<sup>4</sup>

And in February 2016, nearly a hundred years later, there came another equally momentous announcement—this time at the Press Club in Washington—which offered the newest and strongest vindication of Einstein’s theory. This was the detection of gravitational waves by LIGO (the acronym stands for Laser Interferometer Gravitational-Wave Observatory). Einstein envisaged the force of gravity as a “warping” of space. When gravitating objects change their shapes, they generate ripples in space itself. When such a ripple passes the Earth, our local space “jitters”: it is alternately stretched and compressed as gravitational waves pass through it. But the effect is minuscule. This is basically because gravity is such a weak force. The gravitational pull between everyday objects is tiny. If you wave around two dumbbells you will emit gravitational waves—but with quite infinitesimal power. Even planets orbiting stars, or pairs of stars orbiting each other, do not emit at a detectable level.



## **In February 2016 there came another equally momentous announcement: the detection of gravitational waves by the Laser Interferometer Gravitational-Wave Observatory (LIGO), which offered the newest and strongest vindication of Einstein’s theory**

Astronomers are agreed that the sources that LIGO might detect must involve much stronger gravity than in ordinary stars and planets. The best bet is that the events involve black holes or neutron stars. If two black holes form a binary system, they would gradually spiral together. As they get closer, the space around them gets more distorted, until they coalesce into a single, spinning hole. This hole sloshes and “rings,” generating further waves until it settles down as a single quiescent black hole. It is this “chirp”—a shaking of space that speeds up and strengthens until the merger, and then dies away—that LIGO can detect. These cataclysms happen less than once in a million years in our Galaxy. But such an event would give a detectable LIGO signal even if it happened a billion light-years away—and there are millions of galaxies closer than that.

To detect even the most propitious events requires amazingly sensitive—and very expensive—instruments. In the LIGO detectors, intense laser beams are projected along four-kilometer-long pipes and reflected from mirrors at each end. By analyzing the light beams, it is possible to detect changes in the distance between the mirrors, which alternately increases and decreases as “space” expands and contracts. The amplitude of this vibration is exceedingly small, about 0.000000000001 centimeters—millions of times smaller than a single atom. The LIGO project involves two similar detectors about 3,220 kilometers apart—one in Washington State, the other in Louisiana. A single detector would register micro-seismic events, passing vehicles, and so on, and to exclude these false alarms experimenters take note only of events that show up in both.

For years, LIGO detected nothing. But it went through an upgrade, coming fully on line again in September 2015. After literally decades of frustration, the quest succeeded: a “chirp” was detected that signaled the collision of two black holes more than a billion light-years away, and opened up a new field of science—probing the dynamics of space itself.

This detection is, indeed, a big deal: one of the great discoveries of the decade. It allayed any residual skepticism about the validity of Einstein’s equations when LIGO detected events



attributable to a merger of two holes. The detected “chirps,” complicated patterns of oscillations, are excellently fit by computational models based on Einstein’s theory.

The holes detected by LIGO are up to thirty solar masses—the remnants of massive stars. But still more energetic events are expected, involving supermassive holes in the centers of galaxies. When two galaxies merge (as Andromeda and the Milky Way will in about four billion years) the black holes in the center of each will spiral together forming a binary, which will shrink by emitting gravitational radiation and create a strong chirp when the two holes coalesce. Most galaxies have grown via a succession of past mergers and acquisitions. The consequent coalescences of these supermassive black holes would yield gravitational waves of much lower frequencies than ground-based detectors like LIGO can detect. But they are the prime events to which detectors orbiting in space would be sensitive. And ESA has a project called LISA that aims to detect these powerful low-frequency “ripples” in space-time.

### Beyond Galaxies—Cosmic Horizons

We know that galaxies—some disc-like, resembling our Milky Way or Andromeda; others amorphous “ellipticals”—are the basic constituents of our expanding universe. But how much can we actually understand about galaxies? Physicists who study particles can probe them, and crash them together in accelerators at CERN. Astronomers cannot crash real galaxies together. And galaxies change so slowly that in a human lifetime we only see a snapshot of each. But we are no longer helpless: we can do experiments in a “virtual universe” via computer simulations, incorporating gravity and gas dynamics.

We can redo such simulations making different assumptions about the mass of stars and gas in each galaxy, and so forth, and see which matches the data best. Importantly, we find, by this method and others, that all galaxies are held together by the gravity not just of what we see. They are embedded in a swarm of particles that are invisible, but which collectively contribute about five times as much mass as the ordinary atom—the dark matter.

And we can test ideas on how galaxies evolve by observing eras when they were young. Giant telescopes, in space and on the ground, have been used to study “deep fields,” each encompassing a tiny patch of sky. A patch just a few arc minutes across, hugely magnified by these telescopes, reveals hundreds of faint smudges: these are galaxies, some fully the equal of our own, but they are so far away that their light set out more than ten billion years ago. They are being viewed when they have recently formed.

But what happened still further back, before there were galaxies? The key evidence here dates back to 1965, when Penzias and Wilson discovered that intergalactic space is not completely cold. It is warmed to three degrees above absolute zero by weak microwaves, now known to have an almost exact black body spectrum. This is the “afterglow of creation”—the adiabatically cooled and diluted relic of an era when everything was squeezed hot and dense. It is one of several lines of evidence that have allowed us to firm up the “hot big bang” model.

But let us address an issue that might seem puzzling. Our present complex cosmos manifests a huge range of temperature and density—from blazingly hot stars, to the dark night sky. People sometimes worry about how this intricate complexity emerged from an amorphous fireball. It might seem to violate the second law of thermodynamics—which describes an inexorable tendency for patterns and structure to decay or disperse.

The answer to this seeming paradox lies in the force of gravity. Gravity enhances density contrasts rather than wiping them out. Any patch that starts off slightly denser than average



This celestial object looks like a delicate butterfly, but nothing could be farther from reality. What appear to be its wings are roiling cauldrons of gas heated to more than 36,000 degrees Fahrenheit. The gas is tearing across space at more than 600,000 miles an hour—fast enough to travel from Earth to the Moon in twenty-four minutes. The Wide Field Camera 3 (WFC3), a new camera aboard NASA's Hubble Space Telescope, snapped this image in May 2009





would decelerate more, because it feels extra gravity; its expansion lags further and further behind, until it eventually stops expanding and separates out. Many simulations have been made of parts of a “virtual universe”—modeling a domain large enough to make thousands of galaxies. The calculations, when displayed as a movie, clearly display how incipient structures unfold and evolve. Within each galaxy-scale clump, gravity enhances the contrasts still further; gas is pulled in, and compressed into stars.

And there is one very important point. The initial fluctuations fed into the computer models are not arbitrary—they are derived from the variations across the sky in the temperature of the microwave background, which have been beautifully and precisely delineated by ESA’s Planck spacecraft. These calculations, taking account of gravity and gas dynamics, reveal, after the 1000-fold expansion since the photons were last scattered, a cosmos that yields a good statistical fit to the conspicuous present structures and allow the universe’s mean density, age, and expansion rate to be pinned down with a precision of a few percent.

**Many simulations have been made of parts of a “virtual universe”—modeling a domain large enough to make thousands of galaxies. The calculations, when displayed as a movie, clearly display how incipient structures unfold and evolve. Within each galaxy-scale clump, gravity enhances the contrasts still further; gas is pulled in, and compressed into stars**

The fit between the fluctuation spectrum measured by the Planck spacecraft (on angular scales down to a few arc minutes) and a six-parameter model—and the realization that these fluctuations develop, under the action of gravity and gas dynamics, into galaxies and clusters with properties matching our actual cosmos—is an immense triumph. When the history of science in these decades is written, this will be one of the highlights—and I mean one of the highlights of all of science: up there with plate tectonics, the genome, and only very few others.

#### **The Very Early Universe—More Speculative Thoughts**

What about the far future? Any creatures witnessing the Sun’s demise six billion years hence will not be human—they will be as different from us as we are from a bug. Post-human evolution—here on Earth and far beyond—could be as prolonged as the Darwinian evolution that has led to us—and even more wonderful. And, of course, this evolution is even faster now—it happens on a technological timescale, operating far faster than natural selection and driven by advances in genetics and in artificial intelligence (AI). We do not know whether the long-term future lies with organic or silicon-based life.

But what happens even further into the future? In 1998 cosmologists had a big surprise. It was by then well known that the gravity of dark matter dominated that of ordinary stuff—but also that dark matter plus ordinary matter contributed only about thirty percent of the so-called critical density. This was thought to imply that we were in a universe whose expansion was slowing down, but not enough to eventually be halted. But, rather than slowly decelerat-



ing, the redshift versus distance relationship for a particular population of exploding star—Type 1a supernovae—famously revealed that the expansion was speeding up. Gravitational attraction was seemingly overwhelmed by a mysterious new force latent in empty space which pushes galaxies away from each other.

Even after the Sun dies, cosmic expansion could continue. Long-range forecasts are seldom reliable, but the best and most “conservative” bet is that we have almost an eternity ahead—an ever colder and ever emptier cosmos. Galaxies accelerate away and disappear over an “event horizon”—rather like an inside-out version of what happens when things fall into a black hole. All that is left will be the remnants of our Galaxy, Andromeda, and their smaller neighbors. Protons may decay, dark-matter particles annihilate, occasional flashes when black holes evaporate—and then silence.

The nature of dark matter may well be pinned down in a decade, but this dark energy—latent in empty space itself—poses a deeper mystery. It will not be understood until we have a model for the microstructure of space. I am not holding my breath for this: all theorists suspect this will involve phenomena on what is called the “Planck length”—the scale where quantum effects and gravity overlap. This scale is a trillion trillion times smaller than an atom. Dark energy may be the biggest fundamental challenge presented by the present-day universe.

But now, back to the past. The background radiation is a direct messenger of an era when the universe was a few hundred thousand years old—the photons have mainly traveled uninterrupted, without scattering, since then. But we have firm grounds for extrapolating further back—to hotter and denser eras. We are definitely vindicated in extrapolating back to one second, because the calculated proportions of helium and deuterium produced (for a nuclear density fitting other data) match beautifully with what is observed. Indeed, we can probably be confident in extrapolation back to a nanosecond: that is when each particle had about 50 GeV of energy—an energy that can be achieved in the LHC (Large Hadron Collider) at CERN in Geneva—and the entire visible universe was squeezed to the size of our Solar System.

But questions like “Where did the fluctuations come from?” and “Why did the early universe contain the actual mix we observe of protons, photons, and dark matter?” take us back to the even briefer instants when our universe was hugely more compressed still—into an ultra-high-energy domain where experiments offer no direct guide to the relevant physics.

For close to forty years we have had the so-called “inflationary paradigm”—seriously invoking an era when the Hubble radius was a billion times smaller than an atomic nucleus. It is an amazingly bold backward extrapolation, to an era when the physics was extreme, and cannot be tested by experiments. This paradigm is supported already by some evidence. Be that as it may, it might be useful to summarize the essential requirements that must be explained, if we are to understand the emergence of our complex and structured cosmos from simple amorphous beginnings.

1. The first prerequisite is, of course, the existence of the force of gravity—which (as explained earlier) enhances density contrasts as the universe expands, allowing bound structures to condense out from initially small-amplitude irregularities. It is a very weak force. On the atomic scale, it is about forty powers of ten weaker than the electric force between electron and proton. But in any large object, positive and negative charges almost exactly cancel. In contrast, everything has the same “sign” of gravitational charge so when



sufficiently many atoms are packed together, gravity wins. But stars and planets are so big because gravity is weak. Were gravity stronger, objects as large as asteroids (or even sugar-lumps) would be crushed by gravity. So, though gravity is crucial, it is also crucial that it should be very weak.

2. There must be an excess of matter over antimatter.

3. Another requirement for stars, planets, and biospheres is that chemistry should be non-trivial. If hydrogen were the only element, chemistry would be dull. A periodic table of stable elements requires a balance between the two most important forces in the microworld: the nuclear binding force (the “strong interactions”) and the electric repulsive force that drives protons apart.

4. There must be stars—enough ordinary atoms relative to dark matter. (Indeed, there must be at least two generations of stars: one to generate the chemical elements, and a second able to be surrounded by planets.)

5. The universe must expand at the “right” rate—not collapse too soon, nor expand so fast that gravity cannot pull together the structures.

6. There must be some fluctuations for gravity to feed on—sufficient in amplitude to permit the emergence of structures. Otherwise the universe would now be cold ultra-diffuse hydrogen—no stars, no heavy elements, no planets, and no people. In our actual universe, the initial fluctuations in the cosmic curvature have an amplitude 0.00001. According to inflationary models, this amplitude is determined by quantum fluctuations. Its actual value depends on the details of the model.



Another fundamental question is this: how large is physical reality? We can only observe a finite volume. The domain in causal contact with us is bounded by a horizon—a shell around us, delineating the distance light (if never scattered) could have traveled since the big bang. But that shell has no more physical significance than the circle that delineates your horizon if you are in the middle of the ocean. We would expect far more galaxies beyond the horizon. There is no perceptible gradient across the visible universe—suggesting that similar conditions prevail over a domain that stretches thousands of times further. But that is just a minimum. If space stretched far enough, then all combinatorial possibilities would be repeated. Far beyond the horizon, we could all have avatars—and perhaps it would be some comfort that some of them might have made the right decision when we made a wrong one!

But even that immense volume may not be all that exists. “Our” big bang may not be the only one. The physics of the inflation era is still not firm. But some of the options would lead to so-called “eternal inflation” scenario, in which the aftermath of “our” big bang could be just one island of space-time in an unbounded cosmic archipelago.

In scenarios like this, a challenge for twenty-first-century physics is to answer two questions. First, are there many “big bangs” rather than just one? Second—and this is even more interesting—if there are many, are they all governed by the same physics or not? Or is there a huge number of different vacuum states—with different microphysics?

If the answer to this latter question is “yes,” there will still be underlying laws governing the multiverse—maybe a version of string theory. But what we have traditionally called the laws of nature will be just local bylaws in our cosmic patch. Many domains could be stillborn



or sterile: the laws prevailing in them might not allow any kind of complexity. We therefore would not expect to find ourselves in a typical universe—rather, we would be in a typical member of the subset where an observer could evolve. It would then be important to explore the parameter space for all universes, and calculate what domains within it allow complexity to emerge. This cannot be done unless (probably after a long wait) a theory such as string theory becomes believable and “battle-tested.”

Some claim that unobservable entities are not part of science. But few really think that. For instance, we know that galaxies disappear over the horizon as they accelerate away. But (unless we are in some special central position and the universe has an “edge” just beyond the present horizon) there will be some galaxies lying beyond our horizon—and if the cosmic acceleration continues they will remain beyond for ever. Not even the most conservative astronomer would deny that these never-observable galaxies are part of physical reality. These galaxies are part of the aftermath of our big bang. But why should they be accorded higher epistemological status than unobservable objects that are the aftermath of other big bangs?

## **“Our” Big Bang may not be the only one. The physics of the inflation era is still not firm. But some of the options would lead to a so-called “eternal inflation” scenario in which the aftermath of our big bang could be just one island of space-time in an unbounded cosmic archipelago**

To offer an analogy: we cannot observe the interior of black holes, but we believe what Einstein says about what happens there because his theory has gained credibility by agreeing with data in many contexts that we can observe. Likewise, if we had a model that described physics at the energies where inflation is postulated to have occurred, and if that model had been corroborated in other ways, then if it predicts multiple big bangs we should take that prediction seriously.

If there is just one big bang, then we would aspire to pin down why the numbers describing our universe have the values we measure (the numbers in the “standard model” of particle physics, plus those characterizing the geometry of the universe). But if there are many big bangs—eternal inflation, the landscape, and so forth—then physical reality is hugely grander than we would have traditionally envisioned.

It could be that in fifty years we will still be as flummoxed as we are today about the ultra-early universe. But maybe a theory of physics near the “Planck energy” will by then have gained credibility. Maybe it will “predict” a multiverse and in principle determine some of its properties—the probability measures of key parameters, the correlations between them, and so on.

Some do not like the multiverse; it means that we will never have neat explanations for the fundamental numbers, which may in this grander perspective be just environmental accidents. This naturally disappoints ambitious theorists. But our preferences are irrelevant to the way physical reality actually is—so we should surely be open-minded.

Indeed, there is an intellectual and aesthetic upside. If we are in a multiverse, it would imply a fourth and grandest Copernican revolution; we have had the Copernican revolution itself, then the realization that there are billions of planetary systems in our Galaxy; then that there are billions of galaxies in our observable universe.



But we would then realize that not merely is our observable domain a tiny fraction of the aftermath of our big bang, but our big bang is part of an infinite and unimaginably diverse ensemble.

This is speculative physics—but it is physics, not metaphysics. There is hope of firming it up. Further study of the fluctuations in the background radiation will reveal clues. But, more important, if physicists developed a unified theory of strong and electromagnetic forces—and that theory is tested or corroborated in our low-energy world—we would then take seriously what it predicts about an inflationary phase and what the answers to the two questions above actually are.

I started this talk by describing newly discovered planets orbiting other stars. I would like to give a flashback to planetary science four hundred years ago—even before Newton. At that time, Kepler thought that the Solar System was unique, and Earth’s orbit was related to the other planets by beautiful mathematical ratios involving the Platonic regular solids. We now realize that there are billions of stars, each with planetary systems. Earth’s orbit is special only insofar as it is in the range of radii and eccentricities compatible with life (for example, not too cold and not too hot to allow liquid water to exist).

Maybe we are due for an analogous conceptual shift, on a far grander scale. Our big bang may not be unique, any more than planetary systems are. Its parameters may be “environmental accidents,” like the details of the Earth’s orbit. The hope for neat explanations in cosmology may be as vain as Kepler’s numerological quest.

If there is a multiverse, it will take our Copernican demotion one stage further—our Solar System is one of billions of planetary systems in our Galaxy, which is one of billions of galaxies accessible to our telescopes—but this entire panorama may be a tiny part of the aftermath of “our” big bang—which itself may be one among billions. It may disappoint some physicists if some of the key numbers they are trying to explain turn out to be mere environmental contingencies—no more “fundamental” than the parameters of the Earth’s orbit round the Sun. But, in compensation, we would realize that space and time were richly textured—but on scales so vast that astronomers are not directly aware of it, any more than a plankton whose “universe” was a spoonful of water would be aware of the world’s topography and biosphere.

We have made astonishing progress. Fifty years ago, cosmologists did not know if there was a big bang. Now, we can draw quite precise inferences back to a nanosecond. So, in fifty years, debates that now seem flaky speculation may have been firmed up. But it is important to emphasize that progress will continue to depend, as it has up till now, ninety-five percent on advancing instruments and technology—less than five percent on armchair theory, but that theory will be augmented by artificial intelligence and the ability to make simulations.

### Concluding Perspective

Finally, I want to draw back from the cosmos—even from what may be a vast array of cosmoses, governed by quite different laws—and focus back closer to the here and now. I am often asked: is there a special perspective that astronomers can offer to science and philosophy? We view our home planet in a vast cosmic context. And in coming decades we will know whether there is life out there. But, more significantly, astronomers can offer an awareness of an immense future.

Darwinism tells us how our present biosphere is the outcome of more than four billion years of evolution. But most people, while accepting our emergence via natural selection, still

somehow think we humans are necessarily the culmination of the evolutionary tree. That hardly seems credible to an astronomer—indeed, we are probably still nearer the beginning than the end. Our Sun formed 4.5 billion years ago, but it has got six billion more before the fuel runs out. It then flares up, engulfing the inner planets. And the expanding universe will continue—perhaps for ever—destined to become ever colder, ever emptier.

But my final thought is this. Even in this “concertinaed” timeline—extending billions of years into the future, as well as into the past—this century may be a defining moment. Over most of history, threats to humanity have come from nature—disease, earthquakes, floods, and so forth. But this century is special. It is the first where one species—ours—has Earth’s future in its hands, and could jeopardize life’s immense potential. We have entered a geological era called the anthropocene.

Our Earth, this “pale blue dot” in the cosmos, is a special place. It may be a unique place. And we are its stewards at a specially crucial era. That is an important message for us all, whether we are interested in astronomy or not.



## Notes

1. Carl Sagan, *Pale Blue Dot: A Vision of a Human Future in Space*, New York: Random House, 1994.
2. See, for instance, Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, *Gravitation*, New Jersey: Princeton University Press, 2017 (1st ed. 1973).
3. Subrahmanyan Chandrasekhar, *The Mathematical Theory of Black Holes*, Oxford: Oxford University Press, 1998 (1st ed. 1983).
4. *The New York Times*, November 10, 1919.





**José Manuel Sánchez Ron**  
Universidad Autónoma de Madrid

José Manuel Sánchez Ron holds a Bachelor's degree in Physical Sciences from the Universidad Complutense de Madrid (1971) and a PhD in Physics from the University of London (1978). He is a senior Professor of the History of Science at the Universidad Autónoma de Madrid, where he was previously titular Professor of Theoretical Physics. He has been a member of the Real Academia Española since 2003 and a member of the International Academy of the History of Science in Paris since 2006. He is the author of over four hundred publications, of which forty-five are books, including *El mundo después de la revolución. La física de la segunda mitad del siglo XX* (Pasado & Presente, 2014) for which he received Spain's National Literary Prize in the Essay category. In 2011 he received the Jovellanos International Essay Prize for *La Nueva Ilustración: Ciencia, tecnología y humanidades en un mundo interdisciplinar* (Ediciones Nobel, 2011), and, in 2016, the Julián Marías Prize for a scientific career in humanities from the Municipality of Madrid.

Recommended book: *El mundo después de la revolución. La física de la segunda mitad del siglo XX*, José Manuel Sánchez Ron, Pasado & Presente, 2014.

In recent years, although physics has not experienced the sort of revolutions that took place during the first quarter of the twentieth century, the seeds planted at that time are still bearing fruit and continue to engender new developments. This article looks at some of them, beginning with the discovery of the Higgs boson and gravitational radiation. A deeper look reveals the additional need to address other discoveries where physics reveals its unity with astrophysics and cosmology. These include dark matter, black holes, and multiple universes. String theory and supersymmetry are also considered, as is quantum entanglement and its uses in the area of secure communications (quantum cryptography). The article concludes with a look at the presence and importance of physics in a scientifically interdisciplinary world.



Physics is considered the queen of twentieth-century science, and rightly so, as that century was marked by two revolutions that drastically modified its foundations and ushered in profound socioeconomic changes: the special and general theories of relativity (Albert Einstein, 1905, 1915) and quantum physics, which, unlike relativity, cannot be attributed to a single figure as it emerged from the combined efforts of a large group of scientists. Now, we know that revolutions, whether in science, politics, or customs, have long-range effects that may not be as radical as those that led to the initial break, but can nonetheless lead to later developments, discoveries, or ways of understanding reality that were previously inconceivable. That is what happened with physics once the new basic theories were completed. In the case of quantum physics, we are referring to quantum mechanics (Werner Heisenberg, 1925; Paul Dirac, 1925; Erwin Schrödinger, 1926). In Einstein's world, relativistic cosmology rapidly emerged and welcomed as one possible model of the Universe the experimental discovery of the Universe's expansion (Edwin Hubble, 1929). Still, the most prolific "consequences-applications" emerged in the context of quantum physics. In fact, there were so many that it would be no exaggeration to say that they changed the world. There are too many to enumerate here, but it will suffice to mention just a few: the construction of quantum electrodynamics (c. 1949), the invention of the transistor (1947), which could well be called "the atom of globalization and digital society", and the development of particle physics (later called "high-energy physics"), astrophysics, nuclear physics, and solid-state or "condensed matter" physics.

The second half of the twentieth century saw the consolidation of these branches of physics, but we might wonder whether important novelties eventually stopped emerging and everything boiled down to mere developments—what Thomas Kuhn called "normal science" in his 1962 book, *The Structure of Scientific Revolutions*. I hasten to add that the concept of "normal science" is complex and may lead to error: the development of the fundamentals—the "hard core," to use the term introduced by Kuhn—of a scientific paradigm, that is, of "normal science," can open new doors to knowledge of nature and is, therefore, of the greatest importance. In this article, I will discuss the decade between 2008 and 2018, and we will see that this is what has happened in some cases during the second decade of the twenty-first century, significantly after the "revolutionary years" of the early twentieth century.

### The Discovery of the Higgs Boson

One of the most celebrated events in physics during the last decade was the confirmation of a theoretical prediction made almost half a century ago: the existence of the Higgs boson. Let us consider the context that led to this prediction.

High-energy physics underwent an extraordinary advance with the introduction of particles whose names were proposed by one of the scientists responsible for their introduction: Murray Gell-Mann. The existence of these *quarks* was theorized in 1964 by Gell-Mann and George Zweig. Until they appeared, protons and neutrons had been considered truly basic and unbreakable atomic structures whose electric charge was an indivisible unit. Quarks did not obey this rule, as they were assigned fractional charges. According to Gell-Mann and Zweig, hadrons—the particles subject to strong interaction—are made up of two or three types of quarks and antiquarks called *u* (*up*), *d* (*down*) and *s* (*strange*), whose electric charges are, respectively,  $2/3$ ,  $1/3$ , and  $1/3$  of an electron's (in fact, there can be two types of hadrons: baryons—protons, neutrons, and hyperions—and mesons, which are particles whose masses have values between those of an electron and a proton). Thus, a proton is made up of two *u* quarks and one *d*, while a neutron consists of two *d* quarks and one *u*. They are, therefore,



composite structures. Since then, other physicists have proposed the existence of three more quarks: *charm* (*c*; 1974), *bottom* (*b*; 1977) and *top* (*t*; 1995). To characterize this variety, quarks are said to have six *flavors*. Moreover, each of these six can be of three types, or *colors*: red, yellow (or green), and blue. Moreover, for each quark there is an antiquark. (Of course, names like these—color, flavor, up, down, and so on—do not represent the reality we normally associate with such concepts, although in some cases they have a certain logic, as is the case with *color*).

Ultimately, quarks have *color* but hadrons do not: they are white. The idea is that only the “white” particles are observable directly in nature. Quarks are not, as they are “confined,” that is, associated to form hadrons. We will never be able to observe a free quark. Now, in order for quarks to remain confined, there must be forces among them that differ considerably from electromagnetic or other forces. As Gell-Mann put it (1995: 200): “Just as the electromagnetic force among electrons is measured by the virtual exchange of photons, quarks are linked to each other by a force that arises from the exchange of other types: gluons (from the word, *glue*) bear that name because they stick quarks together to form observable white objects such as protons and neutrons.”

## Physics is considered the queen of twentieth-century science, and rightly so, as that century was marked by two revolutions that drastically modified its foundations and ushered in profound socioeconomic changes: the special and general theories of relativity and quantum physics

About a decade after the introduction of quarks, a new theory—quantum chromodynamics—emerged to explain why quarks are so strongly confined that they can never escape from the hadronic structures they form. Coined from *chromos*, the Greek word for color, the term *chromodynamics* alludes to the *color* of quarks, while the adjective *quantum* indicates that it meets quantum requirements. Quantum chromodynamics is a theory of elementary particles with color, which is associated with quarks. And, as these are involved with hadrons, which are the particles subject to strong interaction, we can affirm that quantum chromodynamics describes that interaction.

So quantum electrodynamics and quantum chromodynamics function, respectively, as quantum theories of electromagnetic and strong interactions. There was also a theory of weak interactions (those responsible for radioactive processes such as beta radiation, the emission of electrons in nuclear processes), but it had some problems. A more satisfactory quantum theory of weak interaction arrived in 1967 and 1968, when US scientist Steven Weinberg and British-based Pakistani scientist Abdus Salam independently proposed a theory that unifies electromagnetic and weak interactions. Their model included ideas proposed by Sheldon Glashow in 1960. The Nobel Prize for Physics that Weinberg, Salam, and Glashow shared in 1979 reflects this work, especially after one of the predictions of their theory—the existence of “weak neutral currents”—was corroborated experimentally in 1973 at CERN, the major European high-energy laboratory.

Electroweak theory unified the description of electromagnetic and weak interactions, but would it be possible to move further along this path of unification and discover a formulation that would also include the strong interaction described by quantum chromodynamics?



CERN's Globe exhibition center in Switzerland on a snowy day. This wooden building was given to CERN in 2004 as a gift from the Swiss Confederation to mark fifty years since the organization's foundation







The answer arrived in 1974, and it was yes. That year, Howard Georgi and Sheldon Glashow introduced the initial ideas that came to be known as Grand Unification Theories (GUT).

The combination of these earlier theories constituted a theoretical framework for understanding what nature is made of, and it turned out to have extraordinary predictive capacities. Accordingly, two ideas were accepted: first, that elementary particles belong to one of two groups—bosons or fermions, depending on whether their spin is whole or fractional (photons are bosons, while electrons are fermions)—that obey two different statistics (ways of “counting” groupings of the same sort of particle). These are the *Bose-Einstein statistic* and the *Fermi-Dirac statistic*. Second, that all of the Universe’s matter is made up of aggregates of three types of elementary particles: electrons and their relatives (particles called muons and taus), neutrinos (electronic, muonic, and tauonic neutrinos), and quarks, as well as the quanta associated with the fields of the four forces that we recognize in nature (remember that in quantum physics the wave-particle duality signifies that a particle can behave like a field and vice versa): the photon for electromagnetic interaction, Z and W particles (gauge bosons) for weak interaction, gluons for strong interaction, and, while gravity has not yet been included in this framework, supposed gravitons for gravitational interaction. The subgroup formed by quantum chromodynamics and electroweak theory (that is, the theoretical system that includes relativist theories and quantum theories of strong, electromagnetic, and weak interactions) is especially powerful, given the balance between predictions and experimental proof. This came to be known as the *Standard Model*, but it had a problem: explaining the origin of the mass of the elementary particles appearing therein called for the existence of a new particle, a boson whose associated field would permeate all space, “braking,” so to speak, particles with mass so that, through their interaction with the Higgs field, they showed their mass (it particularly explains the great mass possessed by W and Z gauge bosons, as well as the idea that photons have no mass because they do not interact with the Higgs boson). The existence of such a boson was predicted, theoretically, in three articles published in 1964—all three in the same volume of *Physical Review Letters*. The first was signed by Peter Higgs (1964a, b), the second by François Englert and Robert Brout (1964), and the third by Gerald Guralnik, Carl Hagen, and Thomas Kibble (1964a). The particle they predicted was called “Higgs boson.”

## One of the most celebrated events in physics during the last decade was the confirmation of a theoretical prediction made almost half a century ago: the existence of the Higgs boson

Detecting this supposed particle called for a particle accelerator capable of reaching sufficiently high temperatures to produce it, and it was not until many years later that such a machine came into existence. Finally, in 1994, CERN approved the construction of the Large Hadron Collider (LHC), which was to be the world’s largest particle accelerator, with a twenty-seven-kilometer ring surrounded by 9,600 magnets of different types. Of these, 1,200 were two-pole superconductors that function at minus 217.3°C, which is even colder than outer space, and is attained with the help of liquid helium. Inside that ring, guided by the magnetic field generated by “an escort” of electromagnets, two beams of protons would be accelerated until they were moving in opposite directions very close to the speed of light. Each of these beams would circulate in its own tube, inside of which an extreme vacuum would be maintained,



until it reached the required level of energy, at which point the two beams would be made to collide. The theory was that one of these collisions would produce Higgs bosons. The most serious problem, however, was that this boson almost immediately breaks down into other particles, so detecting it called for especially sensitive instruments. The detectors designed and constructed for the LHC are called ATLAS, CMS, ALICE, and LHCb, and are towering monuments to the most advanced technology.

Following construction, the LHC was first tested by circulating a proton beam on September 10, 2008. The first proton collisions were produced on March 30, 2010, producing a total energy of 7-10<sup>12</sup> eV (that is, 7 tera-electron volts; TeV), an energy never before reached by any particle accelerator. Finally, on July 4, 2012, CERN publicly announced that it had detected a particle with an approximate mass of 125-109 eV (or 125-giga-electron volts; GeV) whose properties strongly suggested that it was a Higgs boson (the Standard Model does not predict its mass). This was front-page news on almost all newspapers and news transmissions around the world. Almost half a century after its theoretical prediction, the Higgs boson's existence had been confirmed. It is therefore no surprise that the 2013 Nobel Prize for Physics was awarded to Peter Higgs and François Englert "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider," as the Nobel Foundation's official announcement put it.

Clearly, this confirmation was cause for satisfaction, but there were some who would have preferred a negative outcome—that the Higgs boson had not been found where the theory expected it to be (that is, with the predicted mass). Their argument, and it was a good one, was expressed by US theoretical physicist and proponent Jeremy Bernstein (2012 a, b: 33) shortly before the discovery was announced: "If the LHC confirms the existence of the Higgs boson, it will mark the end of a long chapter of theoretical physics. The story reminds me of that of a French colleague. A certain parameter had been named after him, so it appeared quite frequently in discussions about weak interactions. Finally, that parameter was measured and the model was confirmed by experiments. But when I went to congratulate him, I found him saddened that his parameter would no longer be talked about. If the Higgs boson failed to appear, the situation would become very interesting because we would find ourselves in serious need of inventing a new physics."

Nonetheless, the fact, and triumph, is that the Higgs boson *does* exist, and has been identified. But science is always in motion, and, in February 2013, the LHC stopped operations in order to make adjustments that would allow it to reach 13 TeV. On April 12, 2018, it began its new stage with the corresponding proton-collision tests. This involved seeking unexpected data that reveal the existence of new laws of physics. For the time being, however, we can say that the Standard Model works very well, and that it is one of the greatest achievements in the history of physics, an accomplishment born of collective effort to a far greater degree than quantum mechanics and electrodynamics, let alone special and general relativity.

Despite its success, however, the Standard Model is not, and cannot be "the final theory." First of all, it leaves out gravitational interaction, and second, it includes too many parameters that have to be determined experimentally. These are the fundamental, yet always uncomfortable whys. "Why do the fundamental particles we detect even exist? Why are there four fundamental interactions, rather than three, five or just one? And why do these interactions exhibit the properties (such as intensity and range of action) they do?" In the August 2011 issue of the American Physical Society's review, *Physics Today*, Steven Weinberg (2011: 33) reflected upon some of these points, and others:



Of course, long before the discovery of neutrino masses, we knew of something else beyond the standard model that suggests new physics at masses a little above 1016 GeV: the existence of gravitation. And there is also the fact that one strong and two electroweak coupling parameters of the standard model, which depends only logarithmically on energy, seem to converge to get a common value at an energy of the order of 1015 GeV to 1016 GeV.

There are lots of good ideas on how to go beyond the standard model, including supersymmetry and what used to be called string theory, but no experimental data yet to confirm any of them. Even if governments are generous to particle physics to a degree beyond our wildest dreams, we may never be able to build accelerators that can reach energies such as 1015 to 1016 GeV. Some day we may be able to detect high-frequency gravitational waves emitted during the era of inflation in the very early universe, that can tell us about physical processes at very high energy. In the meanwhile, we can hope that the LHC and its successors will provide the clues we so desperately need in order to go beyond the successes of the past 100 years.

Weinberg then asks: “What is all this worth? Do we really need to know why there are three generations of quarks and leptons, or whether nature respects supersymmetry, or what dark matter is? Yes, I think so, because answering this sort of question is the next step in a program of learning how all regularities in nature (everything that is not a historical accident) follow from a few simple laws.”

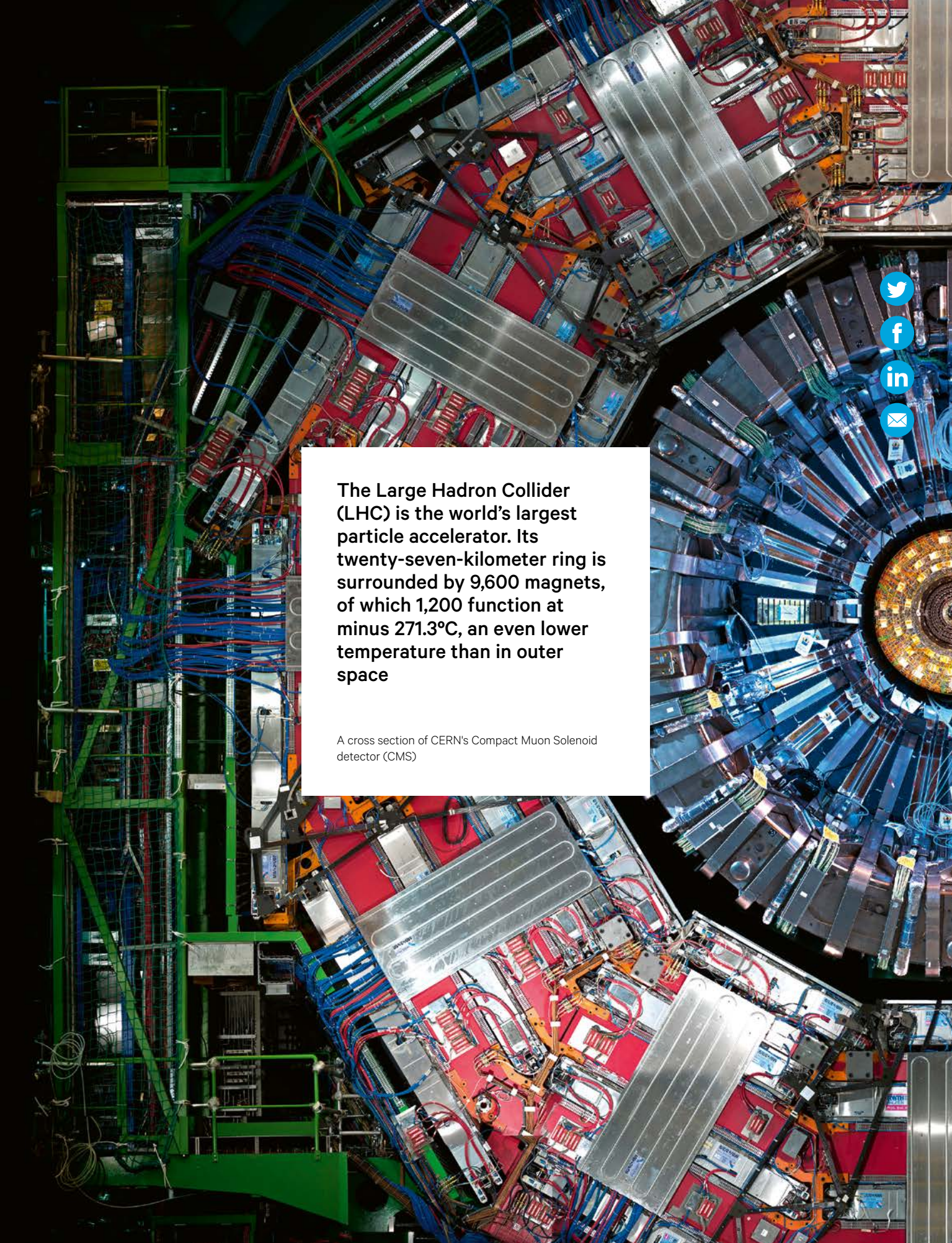
In this quote by Weinberg we see that the energy level at which this “new physics” should clearly manifest, 1015-1016 GeV, is very far from the 13 TeV, that is, the 13·10<sup>3</sup> GeV that the revamped LHC should reach. So far, in fact, that we can perfectly understand Weinberg’s observation that “we may never be able to construct accelerators that can reach those energies.” But Weinberg also pointed out that by investigating the Universe it might be possible to find ways of attaining those levels of energy. He knew this very well, as in the 1970s he was one of the strongest advocates of joining elementary particle physics with cosmology. In that sense, we should remember his book, *The First Three Minutes: A Modern View of the Origin of the Universe* (1977), in which he strove to promote the mutual aid that cosmology and high-energy physics could and in fact *did* obtain by studying the first instants after the *Big Bang*. For high-energy physics that “marriage of convenience” was a breath of fresh air.

Rather than the style and techniques that characterized elementary particle physics in the 1970s, 1980s, and 1990s, Weinberg was referring to something quite different: the physics of gravitational waves or radiation. Besides its other differences, gravitational radiation’s theoretical niche does not lie in quantum physics, but rather in the theory that describes the only interaction that has yet to fit quantum requisites: the general theory of relativity, where the world of basic physics mixes with those of cosmology and astrophysics. And in that plural world, there has also been a fundamental advance over the last decade.

### Gravitational Radiation Exists

Years of intense mental effort began in 1907, with the identification of the so-called “equivalence principle” as a key element for constructing a relativist theory of gravitation. After those years, which included many dead-ends, in November 1915, Albert Einstein completed the structure of what many consider physics’ most elegant theoretical construction: the general theory of relativity. This is a “classic” theory, in the sense that, as I pointed out above, it does not include the principles of quantum theory. And there is consensus about the need for all theories of physics to share those principles. Still, Einstein’s relativist



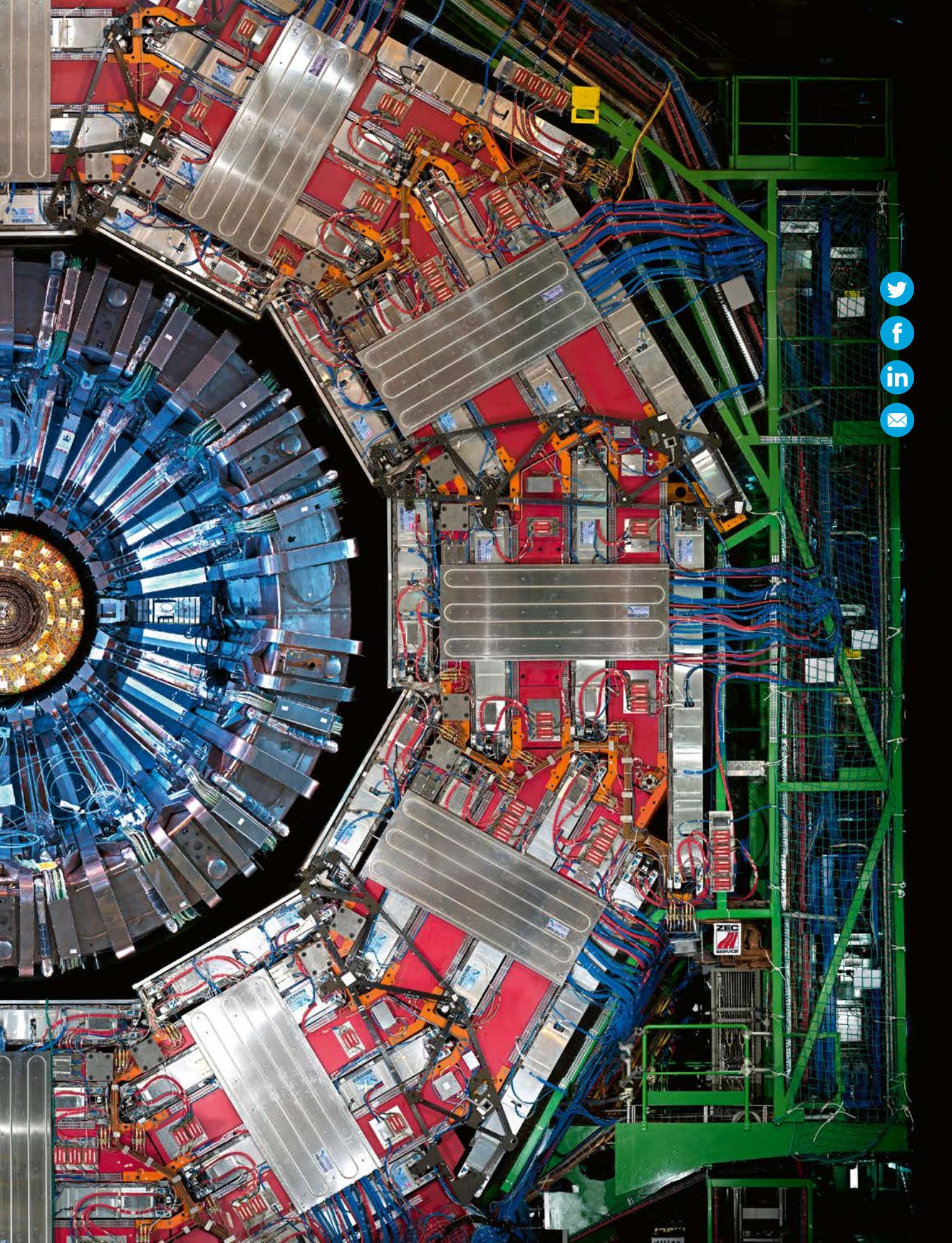


The Large Hadron Collider (LHC) is the world's largest particle accelerator. Its twenty-seven-kilometer ring is surrounded by 9,600 magnets, of which 1,200 function at minus 271.3°C, an even lower temperature than in outer space

A cross section of CERN's Compact Muon Solenoid detector (CMS)











formulation of gravitation has successfully passed every single experimental test yet conceived. The existence of these waves is usually said to have been predicted in 1916, as that is when Einstein published an article concluding that they do, indeed, exist. Still, that work was so limited that Einstein returned to it some years later. In 1936, he and collaborator Nathan Rosen prepared a manuscript titled: “Do Gravitational Waves Exist?” in which they concluded that, in fact, they do not. There were, however, errors in that work, and the final published version (Einstein and Rosen, 1937) no longer rejected the possibility of gravitational waves.

The problem of whether they really exist—essentially, the problem of how to detect them—lasted for decades. No one put more time and effort into detecting them than Joseph Weber, from the University of Maryland, who began in 1960. He eventually came to believe that he had managed to detect them, but such was not the case. His experiment used an aluminum cylinder with a diameter of one meter and a weight of 3.5 tons, fitted with piezoelectric quartz devices to detect the possible distortion of the cylinder when gravitational waves passed through it. When we compare this instrument with the one finally used for their detection, we cannot help but admire the enthusiasm and ingenuousness that characterized this scientist, who died in 2000 without knowing whether his lifelong work was correct or not. Such is the world of science, an undertaking in which, barring exceptions, problems are rarely solved by a single scientist, are often wrought with errors, and take a very long time indeed.

## **On February 11, 2016, a LIGO representative announced that they had detected gravitational waves corresponding to the collision of two black holes. This announcement also constituted a new confirmation of the existence of those singular cosmic entities**

The detection of gravitational waves, which required detecting distortion so small that it is equivalent to a small fraction of an atom, finally occurred in the last decade, when B. P. Abbott (Abbott, et al., 2016) employed a US system called LIGO (Laser Interferometer Gravitational-Wave Observatory) that consisted of two observatories 3,000 kilometers apart (the use of two made it possible to identify false signals produced by local effects), one in Livingston (Louisiana) and the other in Hanford (Washington). The idea was to use interferometric systems with two perpendicular arms in vacuum conditions with an optical path of two or four kilometers for detecting gravitational waves through the minute movements they produce in mirrors as they pass through them. On February 11, 2016, a LIGO representative announced that they had detected gravitational waves and that they corresponded to the collision of two black holes, which thus also constituted a new confirmation of the existence of those singular cosmic entities. While it did not participate in the initial detection (it did not then have the necessary sensitivity, but it was being improved at that time), there is another major interferometric laboratory dedicated to the detection of gravitational radiation: Virgo. Born of a collaboration among six European countries (Italy and France at the helm, followed by Holland, Hungary, Poland, and Spain), it is located near Pisa and has agreements with LIGO. In fact, in the “founding” article, the listing of authors, B. P. Abbott et al., is followed by the statement: “LIGO Scientific Collaboration and Virgo Collaboration.” Virgo soon joined this research with a second round of observations on August 1, 2017.



The detection of gravitational radiation has opened a new window onto the study of the Universe, and it will certainly grow wider as technology improves and more observatories like LIGO and Virgo are established. This is a situation comparable to what occurred in 1930 when Karl Jansky's pioneering radio astronomy (astronomy based on radio waves with wavelengths between a few centimeters and a few meters) experiments radically broadened our knowledge of the cosmos. Before that, our research depended exclusively on the narrow band of wavelengths of the electromagnetic spectrum visible to the human eye. In fact, things have moved quite quickly in this sense. On October 16, 2017, LIGO and Virgo (Abbott et al., 2017) announced that on August 17 of that same year they had detected gravitational radiation from the collision of two neutron stars with masses between 1.17 and 1.60 times the mass of our Sun (remember that neutron stars are extremely dense and small, with a radius of around ten kilometers, comparable to a sort of giant nucleus formed exclusively by neutrons united by the force of gravity). It is particularly interesting that 1.7 seconds after the signal was received, NASA's Fermi space telescope detected gamma rays from the same part of space in which that cosmic collision had occurred. Later, other observatories also detected them. Analysis of this radiation has revealed that the collision of those two stars produced chemical elements such as gold, silver, platinum, and uranium, whose "birthplaces" had been previously unknown.

The detection of gravitational waves also reveals one of the characteristics of what is known as *Big Science*: the article in which their discovery was proclaimed (Abbott et al., 2016) coincided with the LIGO announcement on February 11 and was signed by 1,036 authors from 133 institutions (of its sixteen pages, six are occupied by the list of those authors and institutions).

The importance of LIGO's discovery was recognized in 2017 by the awarding of the Nobel Prize for Physics in two parts. One half of the prize went to Rainer Weiss, who was responsible for the invention and development of the laser interferometry technique employed in the discovery. The other half was shared by Kip Thorne, a theoretical physicist specialized in general relativity who worked alongside Weiss in 1975 to design the project's future guidelines and remains associated with him today; and Barry Barish, who joined the project in 1994 and reorganized it as director. (In 2016, this prize had been awarded to David Thouless, Duncan Haldane, and Michael Kosterlitz, who used techniques drawn from a branch of mathematics known as topology to demonstrate the existence of previously unknown states, or "phases," of matter, for example, superconductors and superfluids, which can exist in thin sheets—something previously considered impossible. They also explained "phase transitions," the mechanism that makes superconductivity disappear at high temperatures.)

## Black Holes and Wormholes

The fact that the gravitational waves first detected in LIGO came from the collision of two black holes also merits attention. At the time, there were already numerous proofs of the existence of these truly surprising astrophysical objects (the first evidence in that sense arrived in 1971, thanks to observations made by instruments installed on a satellite that the United States launched on December 12, 1970, and since then, many more have been identified, including those at the nucleus of numerous galaxies, one of which is our own Milky Way). It is worth remembering that in the 1970s many scientists specializing in general relativity thought that black holes were nothing more than "mathematical ghosts" generated by some solutions to Einstein's theory and thus dismissible. After all, the equations in a physics theory that describes an area of reality can include solutions that do not exist in nature. Such is the case, for example, with relativist cosmology, which includes multiple possible universes. As it happens, black holes *do* exist,



although we have yet to understand such fundamental aspects as where mass goes when they swallow it. The greatest advocates of the idea that black holes are the inevitable consequence of general relativity and must therefore exist were Stephen Hawking and Roger Penrose, who was initially trained in pure mathematics. Their arguments were first proposed in a series of works published in the 1960s, and they were later followed by other scientists, including John A. Wheeler (the director of Thorne's doctoral dissertation) who, in fact, invented the term "black hole." Moreover, before his death on March 14, 2018, Hawking had the satisfaction of recognizing that this new confirmation of the general theory of relativity, to whose development he had dedicated so much effort, also constituted another proof of the existence of black holes. Given that no one—not even Penrose or Wheeler (also now dead)—had contributed as much to the physics of black holes as Hawking, if the Nobel Foundation's statutes allowed its prizes to be awarded to a maximum of four rather than three individuals, he would have been a fine candidate. (In 1973, he presented what is considered his most distinguished contribution: a work maintaining that black holes are not really so "black," because they emit radiation and therefore may end up disappearing, although very slowly. This has yet to be proven.) But history is what it is, not what some of us might wish it were.

## **The greatest advocates of the idea that black holes are the inevitable consequence of general relativity and must therefore exist were Stephen Hawking and Roger Penrose in a series of works published in the 1960s**

In the text quoted above, Weinberg suggests that: "Someday we may be able to detect high-frequency gravitational waves emitted during the very early universe's period of inflation, which could offer data on very high-energy physical processes." This type of gravitational wave has yet to be detected, but it may happen quite soon, because we have already observed the so-called "wrinkles in time," that is, the minute irregularities in the cosmic microwave background that gave rise to the complex structures, such as galaxies, now existing in the Universe (Mather et al., 1990; Smoot et al., 1992). We may recall that this occurred thanks to the *Cosmic Background Explorer* (COBE), a satellite placed in orbit 900 kilometers above the Earth in 1989. Meanwhile, the future of gravitational-radiation astrophysics promises great advances, one of which could be the identification of cosmological entities as surprising as black holes: the "wormholes," whose name was also coined by John Wheeler. Simply put, wormholes are "shortcuts" in the Universe, like bridges that connect different places in it. Take, for example, two points in the Universe that are thirty light-years apart (remember, a light-year is the distance a ray of light can travel in one year). Given the curvature of the Universe, that is, of space-time, there could be a shortcut or bridge between them so that, by following this new path, the distance would be much less: possibly just two light-years, for example. In fact, the possible existence of these cosmological entities arose shortly after Albert Einstein completed the general theory of relativity. In 1916, Viennese physicist Ludwig Flamm found a solution to Einstein's equations in which such space-time "bridges" appeared. Almost no attention was paid to Flamm's work, however, and nineteen years later Einstein and one of his collaborators, Nathan Rosen, published an article that represented physical space as formed by two identical "sheets" connected along a surface they called "bridge." But rather





than thinking in terms of shortcuts in space—an idea that, in their opinion, bordered on the absurd—they interpreted this bridge as a particle.

Decades later, when the general theory of relativity abandoned the timeless realm of mathematics in which it had previously been cloistered, and technological advances revealed the usefulness of understanding the cosmos and its contents, the idea of such shortcuts began to be explored. One result obtained at that time was that, if they indeed exist, it is for a very short period of time—so short, in fact, that you cannot look through them. In terms of travel, they would not exist long enough to serve as a *shortcut* from one point in the Universe to another. In his splendid book, *Black Holes and Time Warps* (1994), Kip Thorne explains this property of wormholes, and he also recalls how, in 1985, he received a call from his friend, Carl Sagan, who was finishing work on a novel that would also become a film, *Contact* (1985). Sagan had little knowledge of general relativity, but he wanted the heroine of his story, astrophysicist Eleanor Arroway (played by Jodie Foster in the film), to travel rapidly from one place in the Universe to another via a black hole. Thorne knew this was impossible, but to help Sagan, he suggested replacing the black hole with a wormhole: “When a friend needs help,” he wrote in his book (Thorne 1994, 1995: 450, 452), “you are willing to look anywhere for it.” Of course, the problem of their very brief lifespan continued to exist, but to solve it, Thorne suggested that Arroway was able to keep the wormhole open for the necessary time by using “an exotic matter” whose characteristics he more-or-less described. According to Thorne, “exotic matter *might* possibly exist.” And in fact, others (among them, Stephen Hawking) had reached the same conclusion, so the question of whether wormholes can be open for a longer period of time than originally thought has led to studies related to ideas that make sense in quantum physics, such as fluctuations in a vacuum: considering space as if, on an ultramicroscopic scale, it were a boiling liquid.

**Simply put, wormholes are “shortcuts” in the Universe, like bridges that connect different places in it. For example, if two points in the Universe are thirty light-years apart, the curvature of the Universe (of space-time) might permit the existence of a shortcut between them—possibly only two light-years in length**

Another possibility recently considered comes from a group of five scientists at Louvain University, the Autonomous University of Madrid-CSIC, and the University of Waterloo, whose article in *Physical Review D* (Bueno, Cano, Goelen, Hertog, and Vernocke, 2018) proposes the possibility that the gravitational radiation detected by LIGO, which was interpreted as coming from the collision of two black holes, might have a very different origin: the collision of two rotating wormholes. Their idea is based on the existence of a border or *event horizon* around black holes, which makes the gravitational waves produced by a collision such as the one detected in 2016 cease in a very short period of time. According to those scientists, this would not happen in the case of wormholes, where such event horizons do not exist. There, the waves should reverberate, producing a sort of “echo.” Such echoes were not detected, but that may be because the instrumentation was unable to do so, or was not prepared for it. This is a problem to be solved in the future.



Seriously considering the existence of wormholes—“bridges” in space-time—might seem like entering a world in which the border between science and science fiction is not at all clear, but the history of science has shown us that nature sometimes proves more surprising than even the most imaginative human mind. So, who really knows whether wormholes might actually exist? After all, before the advent of radio astronomy, no scientist could even imagine the existence of astrophysical structures such as pulsars or quasars. Indeed, the Universe itself, understood as a differentiated entity, could end up losing its most fundamental characteristic: its unicity. Over the last decade scientists have given increasingly serious consideration to a possibility that arose as a way of understanding the collapse of the wave function, the fact that, in quantum mechanics, what finally decides which one of a system’s possible states becomes real (and how probable this occurrence is) is observation itself, as before that observation takes place, all of the system’s states coexist. The possibility of thinking in other terms was presented by a young doctoral candidate in physics named Hugh Everett III. Unlike most of his peers, he was unconvinced by the Copenhagen interpretation of quantum mechanics so favored by the influential Niels Bohr, especially its strange mixture of classical and quantum worlds. The wave function follows its quantum path until subjected to measurement, which belongs to the world of classical physics, at which point it collapses. Everett thought that such a dichotomy between a quantum description and a classical one constituted a “philosophical monstrosity.”<sup>1</sup> He therefore proposed discarding the postulated collapse of the wave function and trying to include the observer in that function.

### **Before the advent of radio astronomy, no scientist could even imagine the existence of astrophysical structures such as pulsars or quasars. Indeed, the Universe itself, understood as a differentiated entity, could end up losing its most fundamental characteristic: its unicity**

It is difficult to express Everett’s theory in just a few words. In fact, John Wheeler, who directed his doctoral dissertation, had trouble accepting all its content and called for various revisions of his initial work, including shortening the first version of his thesis and limiting the forcefulness of some of his assertions, despite the fact that he recognized their value. Here, I will only quote a passage (Everett III, 1957; Barrett and Byrne (eds.), 2012: 188–189) from the article that Everett III (1957) published in *Reviews of Modern Physics*, which coincides with the final version of his doctoral dissertation (successfully defended in April 1957):

We thus arrive at the following picture: Throughout all of a sequence of observation processes there is only one physical system representing the observer, yet there is no single unique *state* of the observer [...] Nevertheless, there is a representation in terms of a *superposition* [...] Thus with each succeeding observation (or interaction), the observer state ‘branches’ into a number of different states. Each branch represents a different outcome of the measurement and the *corresponding* eigenstate for the object-system state. All branches exist simultaneously in the superposition after any given sequence of observations.



In this quote, we encounter what would become the most representative characteristic of Everett's theory. But it was Bryce DeWitt, rather than Everett, who promoted it. In fact, DeWitt recovered and modified Everett's theory, turning it into "the many-worlds interpretation" (or multiverses), in a collection of works by Everett that DeWitt and Neill Graham edited in 1973 with the title *The Many-Worlds Interpretation of Quantum Mechanics* (DeWitt and Graham [eds.], 1973). Earlier, DeWitt (1970) had published an attractive and eventually influential article in *Physics Today* that presented Everett's theory under the provocative title "Quantum Mechanics and Reality." Looking back at that text, DeWitt recalled (DeWitt-Morette, 2011: 95): "The *Physics Today* article was deliberately written in a sensational style. I introduced terminology ('splitting', multiple 'worlds', and so on.) that some people were unable to accept and to which a number of people objected because, if nothing else, it lacked precision." The ideas and the version of Everett's theory implicit in DeWitt's presentation, which have been maintained and have even flourished in recent times, are that the Universe's wave function, which is the only one that really makes sense according to Everett, splits with each "measuring" process, giving rise to worlds and universes that then split into others in an unstoppable and infinite sequence.

## **The ideas and the version of Everett's theory implicit in DeWitt's presentation are that the Universe's wave function splits with each "measuring" process, giving rise to worlds and universes that then split into others in an unstoppable and infinite sequence**

In his *Physics Today* article, DeWitt (1970: 35) wrote that: "No experiment can reveal the existence of 'other worlds.' However, the theory does have the pedagogical merit of bringing most of the fundamental issues of measurement theory clearly into the foreground, hence providing a useful framework for discussion." For a long time (the situation has begun to change in recent years) the idea of multiverses was not taken very seriously, and some even considered it rather ridiculous, but who knows whether it will become possible, at some future time, to imagine an experiment capable of testing the idea that other universes may exist, and, if they do, whether the laws of physics would be the same as in our Universe, or others. Of course, if they were different, how would we identify them?

### **Dark Matter**

Until the end of the twentieth century, scientists thought that, while there was still much to learn about its contents, structure, and dynamics, we knew what the Universe is made of: "ordinary" matter of the sort we constantly see around us, consisting of particles (and radiations/quanta) that are studied by high-energy physics. In fact, that is not the case. A variety of experimental results, such as the internal movement of some galaxies, have demonstrated the existence of matter of an unknown type called "dark matter," as well as something called "dark energy," which is responsible for the Universe expanding even faster than expected. Current results indicate that about five percent of the Universe consists of ordinary mass, twenty-seven percent is dark matter and sixty-eight percent is dark energy.



In other words, we thought we knew about what we call the Universe when, in fact, it is still largely unknown because we have yet to discover what dark matter and dark energy actually are.

At the LHC, there were hopes that a candidate for dark mass particles could be detected. The existence of these WIMPs (weakly interacting massive particles) is predicted by what is called supersymmetry, but the results have so far been negative. One specific experiment that attempted to detect dark matter used a Large Underground Xenon or LUX detector at the Stanford Underground Laboratory and involved the participation of around one hundred scientists and engineers from eighteen institutions in the United States, Europe, and, to a lesser degree, other countries. This laboratory, located 1,510 meters underground in a mine in South Dakota, contains 370 kilos of ultra-pure liquid xenon and the experiment sought to detect the interaction of those particles with it. The results of that experiment, which took place between October 2014 and May 2016, were also negative.

### Supersymmetry and Dark Matter

From a theoretical standpoint, there is a proposed formulation that could include dark matter, that is, the “dark particles” or WIMPs, mentioned above. It consists of a special type of symmetry known as “supersymmetry,” whose most salient characteristic is that for each of the known particles there is a corresponding “supersymmetric companion.” Now, this companion must possess a specific property: its spin must be  $1/2$  less than that of its known partner. In other words, one will have a spin that corresponds to an integer, while the other’s will correspond to a half-integer; thus, one will be a boson (a particle with an integer spin) and the other, a fermion (particles with a semi-integer spin). In that sense, supersymmetry establishes a symmetry between bosons and fermions and therefore imposes that the laws of nature will be the same when bosons are replaced by fermions and vice versa. Supersymmetry was discovered in the early 1970s and was one of the first of a group of theories of other types that raised many hopes for unifying the four interactions—bringing gravitation into the quantum world—and thus moving past the Standard Model. That group of theories is known as string theory.<sup>2</sup> A good summary of supersymmetry was offered by David Gross (2011: 163–164), one of the physicists who stands out for his work in this field:

Perhaps the most important question that faces particle physicists, both theorists and experimentalists, is that of supersymmetry. Supersymmetry is a marvelous theoretical concept. It is a natural, and probably unique, extension of the relativistic and general relativistic symmetries of nature. It is also an essential part of string theory; indeed supersymmetry was first discovered in string theory, and then generalized to quantum field theory. [...]

In supersymmetric theories, for every particle there is a ‘superpartner’ or a ‘superparticle.’ [...] So far, we have observed no superpartners [...] But we understand that this is perhaps not surprising. Supersymmetry could be an exact symmetry of the laws of nature, but spontaneously broken in the ground state of the universe. Many symmetries that exist in nature are spontaneously broken. As long as the scale of supersymmetry breaking is high enough, we would not have seen any of these particles yet. If we observe these particles at the new LHC accelerator then, in fact, we will be discovering new quantum dimensions of space and time.[...]

Supersymmetry has many beautiful features. It unifies by means of symmetry principles fermions, quarks, and leptons (which are the constituents of matter), bosons (which are the quanta of force), the photon, the W, the Z, the gluons in QCD, and the graviton.





Stephen Hawking (1942–2018) aboard a modified Boeing 727 jet owned by the Zero Gravity Corporation. The jet completes a series of steep ascents and dives that create short periods of weightlessness due to free fall. During this flight, Hawking experienced eight such periods





After offering other examples of the virtues of supersymmetry, Gross refers to dark matter: “Finally, supersymmetric extensions of the standard model contain natural candidates for dark-matter WIMPs. These extensions naturally contain, among the supersymmetric partners of ordinary matter, particles that have all the hypothesized properties of dark matter.”

As Gross pointed out, the experiments at the LHC were a good place to find those “supersymmetric dark companions,” which could be light enough to be detected by the CERN accelerator, although even then they would be difficult to detect because they interact neither with electromagnetic force—they do not absorb, reflect, or emit light—nor with strong interaction, because they do not interact with “visible particles,” either. Nonetheless, they possess energy and momentum (otherwise, they would be “ghosts” with no physical entity whatsoever), which opens the doors to inferring their existence by applying the customary laws of conservation of energy-momentum to what is seen after the particles observed collide with the WIMP. All the same, no evidence of their existence has yet been found in the LHC. In any case, the problem of what dark matter actually is constitutes a magnificent example of the confluence of physics (the physics of elemental particles) with cosmology and astrophysics—yet another indication that these fields are sometimes impossible to separate.

### String Theories

The string theories mentioned with regard to supersymmetry appeared before it did. According to string theory, nature’s basic particles are actually unidimensional filaments (extremely thin strings) in space with many more dimensions than the three spatial and one temporal dimension that we are aware of. Rather than saying that they “are” or “consist of” those strings, we should, however, say that they “are manifestations” of the vibrations of those strings. In other words, if our instruments were powerful enough, rather than seeing “points” with certain characteristics we call electron, quark, photon, or neutrino, for example, we would see minute, vibrating strings (whose ends can be open or closed).

The first version of string theory arose in 1968, when Gabriele Veneziano (1968) introduced a string model that appeared to describe the interaction among particles subject to strong interaction. Veneziano’s model only worked for bosons. In other words, it was a theory of bosonic strings, but it did demand a geometrical framework of twenty-six dimensions. It was Pierre Ramond (1971) who first managed—in the work mentioned in footnote 2, which introduced the idea of supersymmetry—to extend Veneziano’s idea to include “fermionic modes of vibration” that “only” required ten-dimensional spaces. Since then, string (or superstring) theory has developed in numerous directions. Its different versions seem to converge to form what is known as M-theory, which has eleven dimensions.<sup>3</sup> In *The Universe in a Nutshell*, Stephen Hawking (2002: 54–57) observed:

I must say that personally, I have been reluctant to believe in extra dimensions. But as I am a positivist, the question ‘do extra dimensions really exist?’ has no meaning. All one can ask is whether mathematical models with extra dimensions provide a good description of the universe. We do not yet have any observations that require extra dimensions for their explanation. However, there is a possibility we may observe them in the Large Hadron Collider in Geneva. But what has convinced many people, including myself, that one should take models with extra dimensions seriously is that there is a web of unexpected relationships, called dualities, between the models. These dualities show that the models are all essentially equivalent; that is, they are just different aspects of the same underlying theory, which has been

given the name M-theory. Not to take this web of dualities as a sign we are on the right track would be a bit like believing that God put fossils into the rocks in order to mislead Darwin about the evolution of life.



Once again we see what high hopes have been placed on the LHC, although, as I already mentioned, they have yet to be satisfied. Of course this does not mean that some string theory capable of bringing gravity into a quantum context may not actually prove true.<sup>4</sup> They are certainly enticing enough to engage the greater public, as can be seen in the success of the one by Hawking mentioned above, or *The Elegant Universe* (1999), written by Brian Greene, another specialist in this field. There are two clearly distinguished groups within the international community of physicists (and mathematicians). Some think that only a version of string theory could eventually provide the possibility of fulfilling the long-awaited dream of unifying the four interactions to form a great quantum synthesis, thus surpassing the Standard Model and general relativity and discovering ways of experimentally proving that theory. Others believe that string theory has received much more attention than it deserves as it is an as yet unprovable formulation more fitting in mathematical settings than in physics (in fact, mathematics has not only strongly contributed to string theories, it has also received a great deal from them. It is hardly by chance that one of the most outstanding string-theory experts, Edward Witten, was awarded the Fields Medals in 1990, an honor considered the mathematics equivalent to the Nobel Prize). As to the future of string theory, it might be best to quote the conclusions of a recent book about them by Joseph Conlon (2016: 235–236), a specialist in that field and professor of Theoretical Physics at Oxford University:

What does the future hold for string theory? As the book has described, in 2015 ‘string theory’ exists as a large number of separate, quasi-autonomous communities. These communities work on a variety of topics range[ing] from pure mathematics to phenomenological chasing of data, and they have different styles and use different approaches. They are in all parts of the world. The subject is done in Philadelphia and in Pyongyang, in Israel and in Iran, by people with all kinds of opinions, appearance, and training. What they have in common is that they draw inspiration, ideas, or techniques from parts of string theory.

It is clear that in the short term this situation will continue. Some of these communities will flourish and grow as they are reinvigorated by new results, either experimental or theoretical. Others will shrink as they exhaust the seam they set out to mine. It is beyond my intelligence to say which ideas will suffer which fate—an unexpected experimental result can drain old subjects and create a new community within weeks.

At this point, Conlon pauses to compare string theory’s mathematical dimension to other physics theories, such as quantum field theory or gravitation, pointing out that “although they may be phrased in the language of physics, in style these problems are far closer to problems in mathematics. The questions are not empirical in nature and do not require experiment to answer.” Many—probably the immense majority of physicists—would not agree.

**According to string theory, nature’s basic particles are actually unidimensional filaments (extremely thin strings) in space with many more dimensions than the three spatial and one temporal dimension that we are aware of. We should, however, say that they “are manifestations” of the vibrations of those strings**



The example drawn from string theory that Conlon next presented was “AdS/CFT Correspondence,” a theoretical formulation published in 1998 by Argentinean physicist Juan Maldacena (1998) that helps, under specific conditions that satisfy what is known as the “holographic principle” (the Universe understood as a sort of holographic projection), to establish correspondence between certain quantum gravity theories and any compatible field or quantum chromodynamic theory. (In 2015, Maldacena’s article was the most frequently mentioned in high-energy physics, with over 10,000 quotes.) According to Conlon: “The validity of AdS/CFT correspondence has been checked a thousand times—but these checks are calculational in nature and are not contingent on experiment.” He continues:

What about this world? Is it because of the surprising correctness and coherence of string theories such as AdS/CFT that many people think string theory is also likely to be a true theory of nature? [...]

Will we ever actually know whether string theory is physically correct? Do the equations of string theory really hold for this universe at the smallest possible scales?

Everyone who has ever interacted with the subject hopes that string theory may one day move forward into the broad sunlit uplands of science, where conjecture and refutation are battled between theorist and experimentalist as if they were ping-pong balls. This may require advances in theory; it probably requires advances in technology; it certainly requires hard work and imagination.

In sum, the future remains open to the great hope for a grand theory that will unify the description of all interactions while simultaneously allowing advances in the knowledge of matter’s most basic structure.

### Entanglement and Cryptography

As we well know, quantum entanglement defies the human imagination. With difficulty, most of us eventually get used to concepts successfully demonstrated by facts, such as indeterminism (Heisenberg’s uncertainty principle of 1927) or the collapse of the wave function (which states, as mentioned above, that we create reality when we observe it; until then, that reality is no more than the set of all possible situations), but it turns out that there are even more. Another of these counterintuitive consequences of quantum physics is entanglement, a concept and term (*Verschränkung* in German) introduced by Erwin Schrödinger in 1935 and also suggested in the famous article that Einstein published with Boris Podolsky and Nathan Rosen that same year. Entanglement is the idea that two parts of a quantum system are in instant “communication” so that actions affecting one will simultaneously affect the other, no matter how far apart. In a letter to Max Born in 1947, Einstein called it “phantasmagorical action at a distance [*spukhafte Fernwirkung*].”

**Quantum entanglement exists, and over the last decade it has been repeatedly proven. One particularly solid demonstration was provided in 2015 by groups from Delft University, the United States National Institute of Standards and Technology, and Vienna University, respectively**





Now, quantum entanglement exists, and over the last decade it has been repeatedly proven. One particularly solid demonstration was provided in 2015 by groups from Delft University, the United States National Institute of Standards and Technology, and Vienna University, respectively. In their article (Herbst, Scheidl, Fink, Handsteiner, Wittmann, Ursin, and Zeilinger, 2015), they demonstrated the entanglement of two previously independent photons 143 kilometers apart, which was the distance between their detectors in Tenerife and La Palma.

In a world like ours, in which communications via Internet and other media penetrate and condition all areas of society, entanglement constitutes a magnificent instrument for making those transmissions secure, thanks to what is known as “quantum cryptography.” The basis for this type of cryptography is a quantum system consisting, for example, of two photons, each of which is sent to a different receptor. Due to entanglement, if one of those receptors changes something, it will instantly affect the other, even when they are not linked in any way. What is particularly relevant for transmission security is the fact that if someone tried to intervene, he or she would have to take some sort of measure, and that would destroy the entanglement, causing detectable anomalies in the system.

Strictly speaking, quantum information exchange is based on what is known as QKD, that is, “Quantum Key Distribution.”<sup>5</sup> What is important about this mechanism is the quantum key that the entangled receptor receives and uses to decipher the message. The two parts of the system share a secret key that they then use to encode and decode messages. Traditional encryption methods are based on algorithms associated with complex mathematical operations that are difficult to decipher but not impossible to intercept. As I have pointed out, such interception is simply impossible with quantum cryptography.

Quantum entanglement augurs the possible creation of a global “quantum Internet.” It is therefore not surprising that recently created companies, such as the Swiss ID Quantique (founded in 2001 as a spinoff of Geneva University’s Applied Physics Group), the US MagiQ, or Australia’s QuintessenceLabs have taken considerable interest in quantum cryptography, as have well-established firms such as HP, IBM, Mitsubishi, NEC, NTT, and Toshiba.

## **In a world like ours, in which communications via Internet and other media penetrate and condition all areas of society, entanglement constitutes a magnificent instrument for making those transmissions secure, thanks to what is known as “quantum cryptography”**

One problem with quantum cryptography is that when classic communications channels, such as optic fiber, are used, the signal is degraded because the photons are absorbed or diffused by the fiber’s molecules (the limit for sending quantum cryptograph messages is around one or two cities). Classic transmission also breaks down over distances, but this can be remedied with relays. That solution does not exist for quantum transmission, however, because, as mentioned above, any intermediate interaction destroys the message’s unity. The best possible “cable” for quantum communication turns out to be the space vacuum, and a significant advance was recently made in that sense. A team led by Jian-Wei Pan, at China’s University of Science and Technology, in Hefei, in collaboration with Anton Zeilinger (one of the greatest specialists in quantum communications and computation) at Vienna Univer-



sity, managed to send quantum messages back and forth between Xinglong and Graz—7,600 kilometers apart—using China’s *Micius* satellite (placed in orbit 500 kilometers from Earth in August 2016) as an intermediary.<sup>6</sup> To broaden its range of action, they used fiber-optic systems to link Graz to Vienna and Xinglong to Beijing. That way, they held a secure seventy-five minute videoconference between the Chinese and Viennese Science Academies using two gigabytes of data, an amount similar to what cellphones required in the 1970s. Such achievements foreshadow the establishment of satellite networks that will allow secure communications of all kinds (telephone calls, e-mail, faxes) in a new world that will later include another instrument intimately related to the quantum phenomena we have already discussed: quantum computing, in which quantum overlap plays a central role.

## **Quantum entanglement augurs the possible creation of a global “quantum Internet.” It is therefore not surprising that well-established firms such as HP, IBM, Mitsubishi, NEC, NTT, and Toshiba have taken considerable interest in quantum cryptography**

While classic computing stores information in bits (0, 1), quantum computation is based on *qubits* (quantum bits). Many physical systems can act as qubits, including photons, electrons, atoms, and Cooper pairs. Contemplated in terms of photons, the principle of quantum overlap means that these can be polarized either horizontally or vertically, but also in combinations lying between those two states. Consequently, you have a greater number of units for processing or storage in a computing machine, and you can carry out various operations at the same time, that is, you can operate in parallel (every bit has to be either 1 or 0, while a qubit can be 1 and 0 at the same time, which allows multiple operations to be carried out at the same time). Obviously, the more qubits a quantum computer employs, the greater its computing capacity will be. The number of qubits needed to surpass computers is believed to be fifty, and IBM recently announced that it had surpassed that threshold... although only for a few nanoseconds. It turns out to be very difficult indeed to keep the qubits entangled (that is, unperturbed) for the necessary period of time. This is extremely complicated because subatomic particles are inherently unstable. Therefore, to avoid what is called “decoherence,” one of the main areas of research in quantum computing involves finding ways to minimize the perturbing effects of light, sound, movement, and temperature. Many quantum computers are being constructed in extremely low-temperature vacuum chambers for exactly that reason, but obviously, industry and governments (especially China, at present) are making significant efforts to advance in the field of quantum computing: Google and NASA, for example, are using a quantum computer built by the Canadian firm, D-Wave Systems, Inc, which is the first to sell this type of machine capable of carrying out certain types of operations 3,600 times faster than the world’s fastest digital supercomputer.

Another possibility associated with quantum entanglement is teletransportation, which Ignacio Cirac (2011: 478) defined as “the transfer of an intact quantum state from one place to another, carried out by a sender who knows neither the state to be teletransported nor the location of the receiver to whom it must be sent.” Teletransportation experiments have already been carried out with photons, ions, and atoms, but there is still a long way to go.



In previous sections, I have focused on developments in what we could call the “most fundamental physics,” but this science is not at all limited to the study of nature’s “final” components, the unification of forces or the application of the most basic principles of quantum physics. Physics is made up of a broad and varied group of fields—condensed matter, low temperatures, nuclear, optical, electromagnetism, fluids, thermodynamics, and so on—and progress has been made in all of them over the last ten years. Moreover, those advances will undoubtedly continue in the future. It would be impossible to mention them all, but I do want to refer to a field to which physics has much to contribute: interdisciplinarity.

We should keep in mind that nature is unified and has no borders; practical necessity has led humans to establish them, creating disciplines that we call physics, chemistry, biology, mathematics, geology, and so on. However, as our knowledge of nature grows, it becomes increasingly necessary to go beyond these borders and unite disciplines. It is essential to form groups of specialists—not necessarily large ones—from different scientific and technical fields but possessed of sufficient general knowledge to be able to understand each other and collaborate to solve new problems whose very nature requires such collaboration. Physics must be a part of such groups. We find examples of interdisciplinarity in almost all settings, one of which would be the processes underlying atmospheric phenomena. These involve all sorts of sciences: energy exchanges and temperature gradients, radiation received from the Sun, chemical reactions, the composition of the atmosphere, the motion of atmospheric and ocean currents, the biology of animals and plants that explains the behavior and reactions of animal and plant species, industrial processes, social modes and mechanisms of transportation, and so on and so on. Architecture and urban studies offer similar evidence. For climate change, energy limitations, atmospheric pollution, and agglomeration in gigantic cities, it is imperative to deepen cooperation among architecture, urban studies, science, and technology without overlooking the need to draw on other disciplines as well, including psychology and sociology, which are linked to the study of human character. We must construct buildings that minimize energy loss, and attempt to approach energy sufficiency, that is, *sustainability*, which has become a catchword in recent years. Fortunately, along with materials with new thermal and acoustic properties, we have elements developed by science and technology, such as solar panels, as well as the possibility of recycling organic waste.

**Nature is unified and has no borders; practical necessity has led humans to establish them, creating disciplines that we call physics, chemistry, biology, mathematics, geology, and so on. However, as our knowledge of nature grows, it becomes increasingly necessary to go beyond those borders and unite disciplines**

One particularly important example in which physics is clearly involved is the “Brain Activity Map Project” which the then president of the United States, Barack Obama, publicly announced on April 2, 2013. In more than one way, this research project is a successor to the great Human Genome Project that managed to *map* the genes composed by our chromosomes. It seeks to study neuron signals and to determine how their flow through neural networks turns into thoughts, feelings, and actions. There can be little doubt as to the importance of



this project, which faces one of the greatest challenges of contemporary science: obtaining a global understanding of the human brain, including its self-awareness. In his presentation, Obama mentioned his hope that this project would also pave the way for the development of technologies essential to combating diseases such as Alzheimer's and Parkinson's, as well as new therapies for a variety of mental illnesses, and advances in the field of artificial intelligence.

It is enough to read the heading of the article in which a group of scientists presented and defended this project to recognize its interdisciplinary nature and the presence of physics therein. Published in the review *Neuron* in 2012, it was titled "The Brain Activity Map Project and the Challenge of Functional Connectomics" and was signed by six scientists: Paul Alivisatos, Miyoung Chun, George Church, Ralph Greenspan, Michael Roukes, and Rafael Yuste (2012).

Equally worthy of note are nanoscience and nanotechnology. The atomic world is one of the meeting places for the natural sciences and technologies based on them. In the final analysis, atoms and the particles they are made of (protons, neutrons, electrons, quarks, and so on) are the "world's building blocks." Until relatively recently, there was not a research field—I am referring to nanotechnology and nanoscience—in which that shared base showed so many potential applications in different disciplines. Those fields of research and development owe their name to a measurement of length, the nanometer (nm), which is one billionth of a meter. Nanotechnoscience includes any branch of science or technology that researches or uses our capacity to control and manipulate matter on scales between 1 and 100 nm. Advances in the field of nanotechnoscience have made it possible to develop nanomaterials and nanodevices that are already being used in a variety of settings. It is, for example, possible to detect and locate cancerous tumors in the body using a solution of 35 nm nanoparticles of gold, as carcinogenic cells possess a protein that reacts to the antibodies that adhere to these nanoparticles, making it possible to locate malignant cells. In fact, medicine is a particularly appropriate field for nanotechnoscience, and this has given rise to nanomedicine. The human fondness for compartmentalization has led this field to frequently be divided into three large areas: nanodiagnostics (the development of image and analysis techniques to detect illnesses in their initial stages), nanotherapy (the search for molecular-level therapies that act directly on affected cells or pathogenic areas), and regenerative medicine (the controlled growth of artificial tissue and organs).

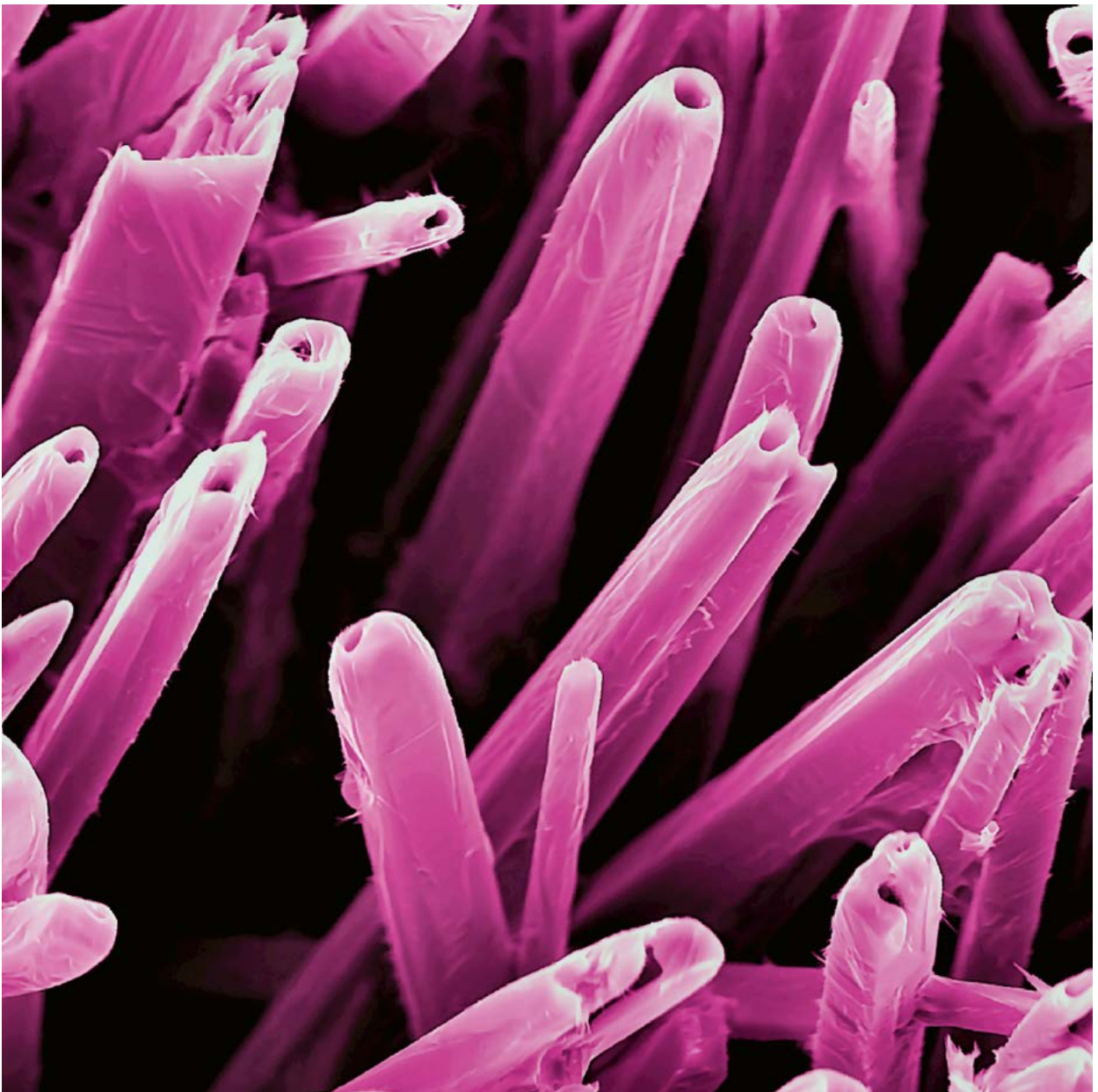
**One particularly important example in which physics is clearly involved is the "Brain Activity Map Project" which the then president of the United States, Barack Obama, publicly announced on April 2, 2013. In more than one way, this research project is a successor to the great Human Genome Project that managed to map the genes composed by our chromosomes**

It is always difficult and risky to predict the future, but I have no doubt that this future will involve all sorts of developments that we would currently consider unimaginable surprises. One of these may well be a possibility that Freeman Dyson (2011), who always enjoys predictions, suggested not long ago. He calls it "radioneurology" and the idea is that, as our knowledge of brain functions expands, we may be able to use millions of microscopic sensors to observe the exchanges among neurons that lead to thoughts, feelings, and so on, and convert them into electromagnetic signals that could be received by another brain outfitted with similar





Colored scanning electron micrograph (SEM) of nanostructures formed on a vanadium and vanadium-oxide surface by a carbon-dioxide laser beam. The resulting nanostructures can have applications in various forms of electronics and nanotechnology



sensors. That second brain would then use them to regenerate the emitting brain's thoughts. Would, or could that become a type of radiotelepathy?



## Epilogue

We live between the past and the future. The present is constantly slipping between our fingers like a fading shadow. The past offers us memories and acquired knowledge, tested or yet to be tested, a priceless treasure that paves our way forward. Of course we do not really know where it will lead us, what new characteristics will appear, or if it will be an easy voyage or not. What is certain is that the future will be different and very interesting. And physics, like all of the other sciences, will play an important and fascinating role in shaping that future.

## Notes

1. Everett to Bryce DeWitt, May 31, 1957; letter reproduced in Barret and Byrne (eds.) (2012: 255).
2. In fact, supersymmetry was actually discovered three times. First by Pierre Ramond, a French physicist at the University of Florida, initially alone, but afterwards in collaboration with André Neveu and John Schwarz; however, the context in which they introduced that new symmetry was very abstract and it was not clear that it had any relation with elementary particles. Around the same time, Yuri Golfand, Evgeny Likhtman, and later Dmitri Volkov and Vladimir Akulov described it in work that never left the Soviet Union. It was work by Julius Wess and Bruno Zumino (1974) that drew the attention of high-energy physicists, although it was considered purely theoretical speculation for quite some time.
3. The idea is that those additional dimensions do not manifest because of a phenomenon known as “compactification.” They exist at subatomic scales and are closed in upon themselves to form circles.
4. Expectations regarding the results that could be obtained at the LHC on both dark matter and string theory have been constant for (too many?) years. One distinguished example in this sense is Edward Witten—about whom I will soon say more—who wrote in his contribution to the book celebrating Stephen Hawking’s sixtieth birthday (Witten 2003: 461): “There is only one major discovery that I can point to as reasonably possible in this decade. That discovery is not theoretical [...] That discovery will probably take place at the Fermilab with the Tevatron [...] or at CERN with the LHC.” The Tevatron ceased activity in 2011.
5. On this point, and others I will discuss below, see Zeilinger (2011).
6. The results were announced in an article published in 2018 and signed by thirty-six researchers, of which twenty-seven were Chinese and nine, Austrian. They belonged to eleven Chinese institutions and three Austrian ones: Sheng-Kai Liao et al. (2018).

## Select Bibliography

- Abbott, B. P. et al. 2016. “Observation of gravitational waves from a binary black hole merger.” *Physical Review Letters* 116: 061102.
- Abbott, B. P. et al. 2017. “Observation of gravitational waves from a binary neutron star inspiral.” *Physical Review Letters* 119: 161101.
- Alivisatos, Paul, Chun, Miyoung, Church, George, Greenspan, Ralph, Roukes, Michael, and Yuste, Rafael. 2012. “The brain activity map project and the challenge of functional connectomics.” *Neuron* 74: 970–974.
- Barrett, Jeffrey A., and Byrne, Peter (eds.). 2012. *The Everett Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- Bernstein, Jeremy. 2012a. “A palette of particles.” *American Scientist* 100(2): 146–155. Reproduced in Bernstein (2012b).
- Bernstein, Jeremy. 2012b. “Un abanico de partículas.” *Investigación y Ciencia* 432: 24–33.
- Bueno, Pablo, Cano, Pablo A., Goelen, Frederik, Hertog, Thomas, and Vernocke, Bert. 2018. “Echoes of Kerr-like wormholes.” *Physical Review D* 7: 024040.
- Chiao, Raymond Y., Cohen, Marvin L., Legget, Anthony J., Phillips, William D., and Harper, Jr., Charles L. (eds.). 2011. *Visions of Discovery*. Cambridge: Cambridge University Press.
- Cirac, J. Ignacio. 2011. “Quantum information.” In Chiao et al. (eds.), 471–495.
- Conlon, Joseph. 2016. *Why String Theory?* Boca Raton: CRC Press.
- DeWitt, Bryce. 1970. “Quantum mechanics and reality.” *Physics Today* 23: 30–35.
- DeWitt, Bryce, and Graham, Neill (eds.). 1973. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- DeWitt-Morette, Cécile. 2011. *The Pursuit of Quantum Gravity. Memoirs of Bryce DeWitt from 1946 to 2004*. Heidelberg: Springer.
- Dyson, Freeman. 2011. “The future of science.” In Chiao et al. (eds.), 39–54.
- Einstein, Albert, and Rosen, Nathan. 1937. “On gravitational waves.” *Journal of the Franklin Institute* 223: 43–54.
- Englert, François, and Brout, Robert. 1964. “Broken symmetry and the mass of gauge vector mesons.” *Physical Review Letters* 13: 321–323.
- Everett III, Hugh. 1957. “Relative state’ formulation of quantum mechanics.” *Reviews of Modern Physics* 29: 454–462.
- Gell-Mann, Murray. 1995. *El quark and el jaguar*. Barcelona: Tusquets (original English: *The Quark and the Jaguar*, 1994).
- Greene, Brian. 1999. *The Elegant Universe*. New York: W. W. Norton.
- Gross, David J. 2011. “The major unknowns in particle physics and cosmology.” In Chiao et al. (eds.), 152–170.
- Guralnik, Gerald S., Hagen, Carl R., and Kibble, Thomas W. 1964. “Global conservation laws and massless particles.” *Physical Review Letters* 13: 585–587.
- Hawking, Stephen. 2002. *El universo en una cáscara de nuez*. Barcelona: Crítica (original English: *The Universe in a Nutshell*, 2001).
- Herbst, Thomas, Scheidl, Thomas, Fink, Matthias, Handsteiner, Johannes, Wittmann, Bernhard, Ursin, Rupert, and Zeilinger, Anton. 2015. “Teleportation of entanglement over 143 km.” *Proceedings of the National Academy of Sciences* 112: 14202–14205.
- Higgs, Peter W. 1964a. “Broken symmetries, massless particles and gauge fields.” *Physics Review Letters* 12: 132–201.
- Higgs, Peter W. 1964b. “Broken symmetries and the masses of gauge bosons.” *Physical Review Letters* 13: 508–509.
- Maldacena, Juan M. 1998. “The large N limit of superconformal field theories and supergravity.” *Advances in Theoretical and Mathematical Physics* 2: 231–252.
- Mather, John C. et al. 1990. “A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite.” *Astrophysical Journal* 354: L-37-L40.
- Ramond, Pierre. 1971. “Dual theory for free fermions.” *Physical Review D* 3: 2415–2418.
- Sheng-Kai Liao et al. 2018. “Satellite-relayed intercontinental quantum network.” *Physical Review Letters* 120: 030501.
- Smoot, George et al. 1992. “Structure of the COBE differential microwave radiometer first year maps.” *Astrophysical Journal* 396: L1-L5.
- Thorne, Kip S. 1994. *Black Holes and Time Warps*. New York: W. W. Norton.
- Thorne, Kip S. 1995. *Agujeros negros y tiempo curvo*. Barcelona: Crítica.
- Veneziano, Gabriele. 1968. “Construction of a crossing-symmetric, Regge-behaved amplitude for linearly rising trajectories.” *Nuovo Cimento A* 57: 190–197.
- Weinberg, Steven. 2011. “Particle physics, from Rutherford to the LHC.” *Physics Today* 64(8): 29–33.
- Wess, Julius, and Zumino, Bruno. 1974. “Supergauge transformations in four dimensions.” *Physics Letters B* 70: 39–50.
- Witten, Edward. 2003. “The past and future of string theory.” In *The Future of Theoretical Physics and Cosmology. Celebrating Stephen Hawking’s 60th Birthday*, G. W. Gibbons, E. P. S. Shellard, and S. J. Rankin (eds.). Cambridge: Cambridge University Press, 455–462.
- Zeilinger, Anton. 2011. “Quantum entanglement: from fundamental questions to quantum communication and quantum computation and back.” In Chiao et al. (eds.), 558–571.



**María Martínón-Torres**  
CENIEH (National Research  
Center on Human Evolution),  
Burgos, Spain

María Martínón-Torres, PhD in Medicine and Surgery, MSc in Forensic Anthropology and MPhil in Human Origins, is currently Director of the National Research Center on Human Evolution (CENIEH) in Burgos (Spain) and Honorary Reader at the Anthropology Department of University College London. Member of the Atapuerca Research Team since 1998 and Research Leader of the Dental Anthropology Group from 2007 to 2015 at CENIEH, she has research interests in hominin palaeobiology, palaeopathology, and the evolutionary scenario of the first Europeans. She has led and participated in several international projects related to the study of the hominin dental evidence worldwide, such as in Dmanisi (Georgia) and China, and published more than sixty book chapters and scientific articles in peer-reviewed journals included in the Science Citation Index, such as *Nature*, *Science*, *PNAS*, and *Journal of Human Evolution*. Her work has been highlighted as Top 1% of most-cited authors in the field of Social Sciences according to Thomson Reuters Essential Science Indicators.

Recommended books: *Orígenes. El universo, la vida, los humanos*, Carlos Briones, Alberto Fernández Soto, and José María Bermúdez de Castro, Crítica, 2015. *La especie elegida*. Juan Luis Arsuaga, Ignacio Martínez, Temas de Hoy, 1998 (Eng. trans: *The Chosen Species: The Long March of Human Evolution*, Wiley-Blackwell, 2005).

Over the last decade, the analysis of ancient DNA has emerged as cutting-edge research that uses methods (genetics) and concepts (hybridization) not previously common in the field of anthropology. Today, we are the only human species on the planet, but we now know that we had offspring with others that no longer exist and have inherited some of their genes. What does it mean to be a hybrid? What are the implications of having the genetic material of other hominins in our blood? Was this hybridization a factor in the Neanderthals' extinction? Does it shift our perspective on human diversity today? Both genetic and fossil evidence gathered over the last decade offer a more diverse and dynamic image of our origins. Many of the keys to *Homo sapiens'* success at adapting may possibly lie in this miscegenation that not only does not harm our identity, but probably constitutes a part of our species' hallmark and idiosyncrasies.





In normal use, the word “investigation” is filled with connotations of progress and future. Paradoxically, its etymology lies in the Latin word *investigare*, which comes, in turn, from *vestigium* (vestige or footprint). Thus, the original meaning of “investigate” would be “to track.” For those of us who study ancient eras, returning to the root of the word “investigate” brings out the frequently overlooked idea that progress is only possible through knowing and learning from the past. Paleoanthropology is the field that studies the origin and evolution of man and tries to reconstruct the history of biological and cultural changes experienced by our ancestors since the lines that have led to humans and chimpanzees split some six million years ago. One of the main bodies of evidence on which the study of human evolution draws is fossils of extinct hominid species. This frequently leads to the erroneous idea that paleoanthropology is an area of study cloistered in the past, and that its contribution to our understanding of today’s human beings consists, at most, of a few anecdotes. In fact, research into human evolution over the last decade has invalidated that paradigm in both the methodological and conceptual sense with research on the very horizon of knowledge that has contributed previously unknown knowledge about our own species.

The past is now uncovered using technology of the future. The need “to make the dead talk” and to maximize the information that can be extracted from cherished and rare fossil and archeological finds has led paleontologists and archeologists to perfect and fully exploit current methods—sometimes to define new lines of investigation. The application, for example, of high-resolution imaging techniques to the study of fossils has spawned an independent methodological branch known as virtual anthropology. Thus, the digital age has also reached the world of the past, and, with these techniques, it is now possible to study a fossil in a nondestructive manner by making two- and three-dimensional measurements and reconstructions of any of its surfaces, whether internal or external. The best example of this fruitful relation between technology and the study of the past, however, appears in the consolidation of a recent research area: paleogenetics, that is, the analysis of ancient DNA. The awarding of the Princess of Asturias Prize for Scientific and Technical Research to Swedish biologist Svante Pääbo, considered one of the founders of paleogenetics, illustrates the importance that molecular studies have attained in the reconstruction of a fundamental part of our history. The team led by Pääbo, director of the Max Planck Institute for Evolutionary Anthropology in Leipzig (Germany), has pioneered the application of high-performance DNA sequencing techniques to the study of ancient DNA, thus making it possible to analyze the complete genomes of extinct organisms. Their collaboration with scientists researching the Sierra de Atapuerca archeological sites in the province of Burgos has, in fact, led to a scientific landmark: extracting DNA from the hominins at the Sima de los Huesos site (430,000 years old). This is the oldest DNA yet recovered in settings without permafrost.

The consolidation of paleogenetic studies has provided data on the origin of our species—*Homo sapiens*—and the nature of our interaction with other now-extinct species of hominins. Such knowledge was unimaginable just ten years ago. Until now, the story of *Homo sapiens*’ origin was practically a linear narrative of its emergence at a single location in Africa. Privileged with a variety of advanced physical and intellectual capacities, modern humans would have spread to all of the continents no more than 50,000 years ago. That is the essence of the “Out of Africa” theory, which suggests that in its expansion across the planet *Homo sapiens* would have replaced all archaic human groups without any crossbreeding at all. Molecular analyses have now dismantled that paradigm, revealing that modern humans not only interbred and produced fertile offspring with now-extinct human species such as Neanderthals, but also that the genetic makeup of today’s non-African human population contains between two and four percent of their genes. Moreover, through genetic analysis of

hand bones discovered in a cave in Denisova, in the Altai mountains of Siberia, geneticists have identified a new human population. We have practically no fossil evidence of them, and therefore know very little about their physical appearance, but we do know that they were different from both Neanderthals and modern humans. Colloquially known as *Denisovans*, they must have coexisted and crossbred with our species and with *Homo Neanderthalensis*, as between four and six percent of the genetic makeup of humans currently living in Papua and New Guinea, Australia and Melanesia is of Denisovan origin.



## **The past is now uncovered using technology of the future. The need to maximize the information that can be extracted from fossil and archeological remains has led to the perfection and exploitation of current methods and the development of new lines of research**

Paleogenetics constitutes frontier research in many of the senses described by science historian Thomas S. Kuhn in his work *The Structure of Scientific Revolutions* (1962), as it has made it possible to use methods (genetics) and concepts (hybridization) that were not previously common in paleoanthropology, thus providing unexpected results that bring the ruling paradigm into question. Moreover, it has opened a door to unforeseen dimensions of knowledge. Although today we are the only human species on Earth, we now know that we were not always alone, and that we knew and had offspring with other humans that no longer exist but have left their genetic mark on us. This triggers a degree of intellectual vertigo. What does it mean to be a hybrid? What was it like to know and coexist with a different intelligent human species? What are the implications of having genetic matter from other human species running through our veins? Was this hybridization involved in the extinction of our brothers, the Neanderthals? Does this shed new light on the diversity of humans today?

It is difficult to resist the fascination of exploring the most ancient periods, despite the relative scarcity of available evidence, but over the last decade paleoanthropology has proved a wellspring of information on more recent times, especially with regard to our own species. Genetic data, along with new fossil discoveries and new datings that place our species outside the African continent earlier than estimated by the “Out of Africa” theory, have uncovered a completely unknown past for today’s humans.

In this context, the present article will review some of the main landmarks of human evolution discovered in the last decade, with a particular emphasis on philosopher Martin Heidegger’s affirmation that “the question is the supreme form of knowledge” and the idea that a diagnosis of any science’s health depends not so much on its capacity to provide answers as to generate new questions—in this case, about ourselves.

### **Our Hybrid Origin**

One of the main problems with the idea of hybrid species is that it contradicts the *biological concept of species* originally formulated by evolution biologist Ernst Mayr. According to Mayr, a species is a group or natural population of individuals that can breed among themselves but are reproductively isolated from other similar groups. This concept implies that indi-



viduals from different species should not be able to crossbreed or have fertile offspring with individuals who do not belong to the same taxon. And yet, nature offers us a broad variety of hybrids—mainly in the plant world, where species of mixed ancestry can have offspring that are not necessarily sterile. This is less common among animals or at least less known, although there are known cases among species of dogs, cats, mice, and primates, including howler monkeys. One particular example among primates is the hybrids that result from crossbreeding different species of monkeys from the genus *Cercopithecidae*, familiarly known as baboons. As we shall see, they have provided us with very useful information. An additional problem is that, until molecular techniques were applied to paleontological studies, the concept of biological species was difficult to establish in fossils. We would have needed a time machine to discover whether individuals from different extinct species could (and did) interact sexually, and whether those interactions produced fertile offspring. Moreover, there is very little information about the physical appearance of the hybrids, which further complicates attempts to recognize them in fossil evidence.

In the context of paleontology, the concept of species is employed in a much more pragmatic way, as a useful category for grouping individuals whose anatomical characteristics (and ideally, their behavior and ecological niche) place them in a homogenous group that can, in principle, be recognized and distinguished from other groups. This clear morphological distinction among the individuals we potentially assign to a species is used as an indirect indicator of their isolation from other populations (species) that would not have maintained their characteristic morphology if they had regularly interbred. Does the fact that *H. sapiens* and *H. Neanderthalensis* interbred signify that they are not different species? Not necessarily, although the debate is open.

## Today we are the only human species on the planet, but we now know that we had offspring with others that no longer exist and that we have inherited some of their genes

The Neanderthals are hominins that lived in Europe some 500,000 years ago. Their extinction some 40,000 years ago roughly coincides with the arrival of modern humans in Europe, and this has led to the idea that our species may have played some role in their disappearance. Few species of the genus *Homo* have created such excitement as the Neanderthals, due to their proximity to our species and their fateful demise. The study of skeletal and archeological remains attributed to them offers an image of a human whose brain was of equal or even slightly larger size than ours, with proven intellectual and physical capacities. They were skilled hunters, but also expert in the use of plants, not only as food but also for medicinal purposes. Moreover, the Neanderthals buried their dead and used ornaments, and while their symbolic and artistic expression seems less explosive than that of modern humans, it does not differ substantially from that of the *H. sapiens* of their time. Indeed, the sempiternal debate as to the possible cultural “superiority” of *H. sapiens* to *H. Neanderthalensis* often includes an erroneous comparison of the latter’s artistic production to that of modern-day humans, rather than to their contemporaneous work. The recent datings of paintings in various Spanish caves in Cantabria, Extremadura, and Andalusia indicate they are earlier than the arrival of modern humans in Europe, which suggests that the Neanderthals may have been responsible for this type of art. If they were thus sophisticated and similar to us in their

complex behavior, and able to interbreed with us, can we really consider them a different species? This debate is further heightened by genetic studies published in late 2018 by Viviane Slon and her colleagues, which suggest that Neanderthals, modern humans, and Denisovans interbred quite frequently. This brings into question the premise that some sort of barrier (biological, cultural, or geographic) existed that hindered reproductive activity among two supposedly different species.

Fossil and genetic data now available are still insufficient to support any firm conclusions, but it is important to point out that even when we are speaking of relatively *frequent* crossbreeding among individuals from different species (and living species have already supplied us with examples of this) we are not necessarily saying that such crosses are the norm in such a group's biological and demographic history. There are circumstances that favor such crossbreeding, including periods in which a given population suffers some sort of demographic weakness, or some part of that population, perhaps a marginal part, encounters difficulty in procreating within its own group; or mainly in periods of ecological transition from one ecosystem to another where two species with different adaptations coexist. Ernst Mayr himself clarified that the isolation mechanisms that separate one lineage from another in reproductive terms consist of biological properties of individuals that prevent habitual crossbreeding by those groups, and while occasional crossbreeding might occur, the character of that exchange is not significant enough to support the idea that the two species have completely merged. And that is the key element here. The Neanderthals are probably one of the human species most clearly characterized and known through fossil evidence. Their low and elongated crania, with an obvious protuberance at the back ("occipital bun"), the characteristic projection of their faces around the nasal region (technically known as mid-facial prognathism) and the marked bone ridge at the brow (supraorbital ridge) are Neanderthal characteristics that remained virtually intact from their origin almost half a million years ago in Europe through to their extinction. If there really had been systematic and habitual interbreeding among Neanderthals and modern humans, it would probably have attenuated or modified those patterns. But in fact, the most pronounced examples of those morphological traits appear in the late Neanderthals. Moreover, except for genetic analyses, we have found no evidence that the two groups lived together. The digs at Mount Carmel, Israel, are the clearest example of physical proximity among the two species, but, in all cases, evidence of one or the other group appears on staggered layers—never on the same one. Thus, Neanderthals and humans may have lived alongside each other, but not "together," and while there may have been occasional crossbreeding, this was not the norm. Therefore, modern humans cannot be considered a fusion of the two. In this sense, the hypothesis that Neanderthals disappeared because they were absorbed by modern humans loses strength.

A final interesting take on inter-species hybrids emerges from baboon studies by researchers such as Rebecca Ackermann and her team. It is popularly thought that a hybrid will present a morphology halfway between the two parent species, or perhaps a mosaic of features from both. But Ackermann and her colleagues have shown that hybrids frequently resemble *neither of the parents*, so that many of their characteristics are actually "morphological novelties." Hybrids tend to be either clearly smaller or larger than their parents, with a high number of pathologies and anomalies that rarely appear in either of the original populations. Some of these anomalies, such as changes in the craniofacial suture, bilateral dental pathologies, and cranial asymmetries, undoubtedly reflect developmental "mismatches." Thus, even when hybridization is possible, the "fit" can be far from perfect. And, in fact, paleogenomic studies, such as those by Fernando Méndez and his team, address the possibility that modern humans may have developed some kind of immune response to the Neanderthal "Y" chromosome,







**For some experts, hybridization may have been advantageous for our species as a source of genes beneficial to our conquest of the world**

A hamadryade ape (*Papio hamadryas*) from the genus *Cercopithecidae*. Studies of specimens resulting from crossbreeding among different species of this family have provided very useful information for the study of hybridization among hominins





**The recent dating of paintings from various Spanish caves opens the possibility that Neanderthals may have been the authors of this type of artistic expression**

Painting at the Altamira Cave in Cantabria, Spain, which dates from the Upper Paleolithic



which would have repeatedly caused *Homo sapiens* mothers impregnated by Neanderthal fathers to miscarry male fetuses. This would, in turn, have threatened the conservation of Neanderthal genetic material.



## **Paleogenomic studies suggest that modern humans may have developed some kind of immune response to the Neanderthal “Y” chromosome, which would have repeatedly caused *Homo sapiens* mothers impregnated by Neanderthals to miscarry male fetuses**

Along that line, we could consider that while hybridization may not have caused Neanderthals to disappear, it may have been a contributing factor. Far from the classic example of dinosaurs and meteors, extinction in the animal world is generally a slow process in which a delicate alteration of the demographic equilibrium does not require major events or catastrophes to tip the scales one way or the other. While the Neanderthals' fertility may have been weakened by the mix, some experts suggest that by interbreeding with Neanderthals and Denisovans our species may have acquired genetic sequences that helped us adapt to new environments after we left Africa, especially through changes in our immune system. Thus, hybridization would have been advantageous for our species as a source of genes beneficial to our conquest of the world.

Further research is necessary to explore the effect of genetic exchange on the destiny of each of these species. But, while hybridization may have negatively affected Denisovans and/or Neanderthals, as they are the ones that disappeared, paleogenetics suggest that the interaction of modern humans with those extinct species was not necessarily violent. One of the most classic theories proposed to explain the Neanderthals' disappearance is confrontation, possibly even violent conflict, between the two species. *Homo sapiens* has been described as a highly “invasive” species, and its appearance, like Attila's horse, has been associated with the extinction of many species of large animals (“megafaunal extinction”), including the Neanderthals. While sex does not necessarily imply love, the fact that there is a certain amount of Neanderthal DNA running through our veins suggests that someone had to care for and assure the survival of hybrid children, and that reality may allow us to soften the stereotype of violent and overpowering *Homo sapiens*.

### **The “Hard” Evidence of Our Origin**

There is no doubt that technological advances in recent years have led us to examine the small and the molecular. The methodological revolution of genetic analysis is bolstered by the birth of paleoproteomics (the study of ancient proteins), a discipline whose importance can only grow in the coming years. Nonetheless, the hard core and heart of anthropology has been and will continue to be fossils. Without fossils, there would be no DNA and no proteins; we would lack the most lasting and complete source of data on which paleoanthropology draws. While futuristic analysis techniques continue to emerge, reality reminds us that to move forward in this field we must literally and figuratively “keep our feet on the ground.” The ground is where we must dig, and that is where, with hard work and no little luck, we find the bones of our forebears. There is still much ground to be covered, and our maps are filled with huge fossil



gaps. Regions such as the Indian Subcontinent or the Arabian Peninsula have barely been explored in that sense, so discoveries there are like new pieces that oblige us to reassemble the puzzle. Over the last decade, new and old fossils found in Asia are shifting the epicenter of attention toward the Asian continent and there are, foreseeably, many surprises in store. Even with regard to the reconstruction of our own species' history, which has long been told exclusively in terms of Africa, the fossils discovered over the last decade have something to say.

From the standpoint of fossils, the hypothesis of our species' African origin has rested mainly on the discovery there of the oldest remains attributable to *Homo sapiens*. These include the Herto and Omo skulls from Ethiopia, which are between 160,000 and 180,000 years old. The Qafzeh and Skhul sites in the Near East have provided an important collection of fossils also attributed to our species, which are between 90,000 and 120,000 years old. And yet, the "Out of Africa" hypothesis holds that our species was not able to enter Europe and Asia until some 50,000 years ago, so the presence of *Homo sapiens* in Israel was not considered dispersion or "exodus" as such.

Over the last decade, however, a significant number of fossils brings the 50,000-year date into question. These include the teeth and jaw found in Daoxian (Fuyan Cave) and Zhirendong, in South China, and the finger bone discovered in Al-Wusta (Saudi Arabia), which place our species outside Africa at least 80,000 years ago, although their presence may even be earlier than 100,000 years ago. With the discovery at a dig in Misliya (Israel) of a human jawbone dating from around 190,000 years ago—this is as old as the oldest African fossils attributable to *Homo sapiens*—it is becoming increasingly clear that our species was able to adapt to other territories earlier than we thought, although the debate is still open. Arguably, genetic evidence continues to suggest that modern humanity comes mainly from a process of dispersion that took place around 50,000 years ago. That does not, however, rule out earlier forays that may not have left any mark on modern humans—or perhaps we have simply not detected them yet. When we limit ourselves to maps with arrows representing the spread of humans, we may easily forget that hominins do not migrate in linear, directional ways, as if they were on an excursion or march with a predetermined destination or purpose. Like any other animal's, human migration should be understood as the expansion or broadening of an area of occupation by a group when a lack of barriers (ecological or climactic, for example) and the presence of favorable demographic conditions allow them to increase their territory. The "Out of Africa" migration was probably not a single event or voyage, but rather a more or less continuous flow of variable volume. There may have been various "Out of Africa" movements, and also several "Into Africa," reentries that are not technically returns, because hominins do not "go home." Instead, they expanded the diameter of their territory whenever nothing kept them from doing so.

## **Arguably, genetic evidence continues to suggest that modern humanity comes mainly from a process of dispersion that took place around 50,000 years ago**

Finally, 300,000-year-old fossil remains at Jebel Irhoud in Morocco shed new light (new questions?) on our species' origin. While those specimens lack some of the features considered exclusively *Homo sapiens* (such as a chin, vertical forehead or high, bulging cranium), many researchers consider them the oldest representative of our lineage. The novelty lies





The skull and neck vertebrae of an adult human from Sima de los Huesos at the Atapuerca site in Burgos, Spain, discovered in 1984. The teeth, jaw, and facial bones appear to be Neanderthal, while the cranium is primitive. This suggests that the Neanderthals' characteristic features evolved separately, following an evolutionary model known as accretion





not so much in their age as in their location. Until recently, most African fossils attributed to our species were found in regions of East or South Africa, not in the North. This and other fossils finds on that continent lend weight to the hypothesis that we originate from not one, but various populations that came to inhabit far-flung regions of the enormous African continent and maintained intermittent genetic exchanges. *Homo sapiens* would thus have evolved along a more cross-linked and nonlinear path than previously thought. This theory, now known as “African Multiregionalism,” suggests that the deepest roots of *Homo sapiens* are already a mixture, a melting pot of different populations of highly varied physical and cultural characteristics. Curiously, the term “multiregionalism” refers to another of the major theories proposed during the twentieth century to explain *Homo sapiens*’ origin. Unlike “Out of Africa”, “Multiregionalism” maintained that *our* species grew out of the parallel evolution of various lineages in different parts of the world, and this theory also observed genetic exchanges among those parallel groups. Surprisingly, the last decade has seen the two poles (Out of Africa and Multiregionalism) growing ever closer together.

### The Legacy of the Past

Both genetic studies and fossil evidence from the last ten years offer a more diverse, rich, and dynamic view of our own species. From its origin in Africa to hybridization with Neanderthals and Denisovans, *Homo sapiens* emerges as a melting pot of humanities. Many of the keys to our successful adaptation as we conquered ever-wider territories and changing environments may well be the result of precisely the cosmopolitan miscegenation that has characterized us for at least the last 200,000 years. This admixture not only does not weaken our identity as a species; it is probably part and parcel of our idiosyncrasy.

Human evolution rests precisely on biological diversity, an advantageous and versatile body of resources on which nature can draw when circumstances require adaptive flexibility. Endogamic and homogeneous species are more given to damaging mutations, and it is even possible that the Neanderthals’ prolonged isolation in Europe over the course of the Ice Age may have made them more vulnerable in the genetic sense. Part of the flexibility that characterizes us today was received from other humans that no longer exist. We are the present and the future, but we are also the legacy of those who are no longer among us.

**Both genetic studies and fossil evidence from the last ten years offer a more diverse, rich, and dynamic view of our own species. From its origin in Africa to hybridization with Neanderthals and Denisovans, *Homo sapiens* emerges as a melting pot of humanities**

Despite being from species who probably recognized each other as different, humans and others now extinct crossbred, producing offspring and caring for them. This inevitably leads us to reflect on current society and its fondness for establishing borders and marking limits among individuals of the same species that are far more insurmountable than those dictated by biology itself. Our culture and social norms frequently take paths that seem to contradict our genetic legacy. How would we treat another human species today? Why are we the only one that survived? Would there even be room for a different form of humans?

Would there be room for difference? What is our level of tolerance toward biological and cultural diversity?

We continue to evolve. Natural selection continues to function, but we have altered selective pressures. Social pressure now has greater weight than environmental pressure. It is more important to be well connected than to enjoy robust health. With the rise of genetic editing techniques, humans now enjoy a “superpower” that we have yet to really control. Society must therefore engage in a mature and consensual debate about where we want to go, but that debate must also consider our own evolutionary history, including our species’ peculiarities and the keys to our success. By any measure, what made us strong was not uniformity, but diversity. Today, more than ever, humanity holds the key to its own destiny. We have become the genie of our own lamp. We can now make a wish for our future, but we have to decide what we want to wish for. We boast about our intelligence as a species, but what we do from now on will determine how much insight we really possess. In ten, twenty, or one hundred years, our past will speak for us, and it will be that past that issues the true verdict on our intelligence.



## Select Bibliography

- Ackermannn, R. R., Mackay, A., Arnold, L. 2016. "The hybrid origin of 'modern humans.'" *Evolutionary Biology* 43, 1–11.
- Gittelmann, R. M., Schraiber, J. G., Vernot, B., Mikacenic, C., Wurfel, M. M., Akey, J. M. 2016. "Archaic hominin admixture facilitated adaptation to out-of-Africa environments." *Current Biology* 26, 3375–3382.
- Groucutt, H. S., Grün, R., Zalmout, I. A. S., Drake, N. A., Armitage, S. J., Candy, I., Clark-Wilson, R., Louys, R., Breeze, P. S., Duval, M., Buck, L. T., Kivell, T. L. [...]. Petraglia, M. D. 2018. "*Homo sapiens* in Arabia by 85,000 years ago." *Nature Ecology and Evolution* 2, 800–809.
- Hershkovitz, I., Weber, G. W., Quam, R., Duval, M., Grün, R., Kinsley, L., Ayalon, A., Bar-Matthews, M., Valladas, H., Mercier, N., Arsuaga, J. L., Martínón-Torres, M., Bermúdez de Castro, J. M., Fornai, C. [...] 2018. "The earliest modern humans outside Africa." *Science* 359, 456–459.
- Hublin, J. J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., Bergmann, I., Le Cabec, A., Benazzi, S., Harvati, K., Gunz, P. 2017. "New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*." *Nature* 546, 289–292.
- Kuhn, T. S. 2006. *La estructura de las revoluciones científicas*. Madrid: Fondo de Cultura Económica de España.
- Liu, W., Martínón-Torres, M., Cai, Y. J., Xing, S., Tong, H. W., Pei, S. W., Sier, M. J., Wu, X. H., Edwards, L. R., Cheng, H., Li, Y. Y., Yang, X. X., Bermúdez De Castro, J. M., Wu, X. J. 2015. "The earliest unequivocally modern humans in southern China." *Nature* 526, 696–699.
- Martínón-Torres, M., Wu, X., Bermúdez de Castro, J. M., Xing, S., Liu, W. 2017. "*Homo sapiens* in the Eastern Asian Late Pleistocene." *Current Anthropology* 58, S434–S448.
- Méndez, F. L., Poznik, G. D., Castellano, S., Bustamante, C. D. 2016. "The divergence of Neandertal and modern human Y chromosomes." *The American Journal of Human Genetics* 98, 728–735.
- Meyer, M., Arsuaga, J. L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., Gracia, A., Bermúdez de Castro, J. M., Carbonell, E., Viola, B., Kelso, J., Prüfer, K., Pääbo, S. 2016. "Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins." *Nature* 531, 504–507.
- Scerri, E. M. L., Thomas, M. G., Manica, A., Gunz, P., Stock, J. T., Stringer, C., Grove, M., Groucutt, H. S., Timmermann, A., Rightmire, G. P., d'Errico, F., Tryon, C. A. [...] Chikhi, L. 2018. "Did our species evolve in subdivided populations across Africa, and why does it matter?" *Trends in Ecology & Evolution* 33, 582–594.
- Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hadinjak, M., Peyrégne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Kelso, J., Meyer, M., Prüfer, K., Pääbo, S. 2018. "The genome of the offspring of a Neanderthal mother and a Denisovan father." *Nature* 561, 113–116.





**Alex Pentland**  
MIT Media Lab

Professor Alex “Sandy” Pentland directs the MIT Connection Science and Human Dynamics labs and previously helped create and direct the MIT Media Lab and the Media Lab Asia in India. He is one of the globally most-cited computational scientists, with Forbes declaring him one of the “seven most powerful data scientists in the world” and he is a founding member of advisory boards for Google, AT&T, Nissan, and the UN Secretary General, a serial entrepreneur who has cofounded more than a dozen companies. He is a member of the US National Academy of Engineering and leader within the World Economic Forum. His most recent books are *Social Physics* and *Honest Signals*.

Recommended book: *Social Physics*, Alex Pentland, Penguin, 2015.

**Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time, designing, monitoring, and evaluating effective policies become almost impossible. Today there are unprecedented possibilities for informing and transforming society and protecting the environment. I describe the social science and computer architecture that will allow this data to be safely used to help adaptation to the new world of data, a world that is more fair, efficient, and inclusive, and that provides greater opportunities than ever before.**



For me, the story starts more than twenty years ago when I was exploring wearable computing. My lab had the world's first cyborg group: about twenty students who soldered together PCs and motorcycle batteries and little lasers so you could shoot images right into your eye. We tried to experiment with what the future would be. To put this in context, remember that there were very few cellphones in 1996. There was not even WiFi. Computers were big, hot things that sat on desks or in air-conditioned rooms. It was clear to me, however, that computers were going to be on our bodies and then basically everywhere.

We built this set of equipment for a whole group of people and then did lots of experiments with it. One of the first things everybody said about it was: "This is really cool, but I'll never wear that."

So, my next step was to get fashion schools involved in our work. The image on the facing page is from a French fashion school called Creapole. I had talked with students about where the technology was going and then they came up with designs. Intriguingly, they essentially invented things that looked like Google Glass and an iPhone down to the fingerprint reading.

It is interesting that by living with prototypes of the technology you can begin to see the future perhaps better than just by imagining it. The truly important thing that we learned by living this future was that vast amounts of data would be generated. When everybody has devices on them, when every interaction is measured by digital sensors, and when every device puts off data you can get a picture of society that was unimaginable a few years ago. This ability to see human behavior continuously and quantitatively has kicked off a new science (according to articles in *Nature*<sup>1</sup> and *Science*<sup>2</sup> and other leading journals) called computational social science, a science which is beginning to transform traditional social sciences. This is akin to when Dutch lens makers created the first practical lenses: microscopes and telescopes opened up broad new scientific vistas. Today, the new technology of living labs—observing human behavior by collecting a community's digital breadcrumbs—is beginning to give researchers a more complete view of life in all its complexity. This, I believe, is the future of social science.

### A New Understanding of Human Nature

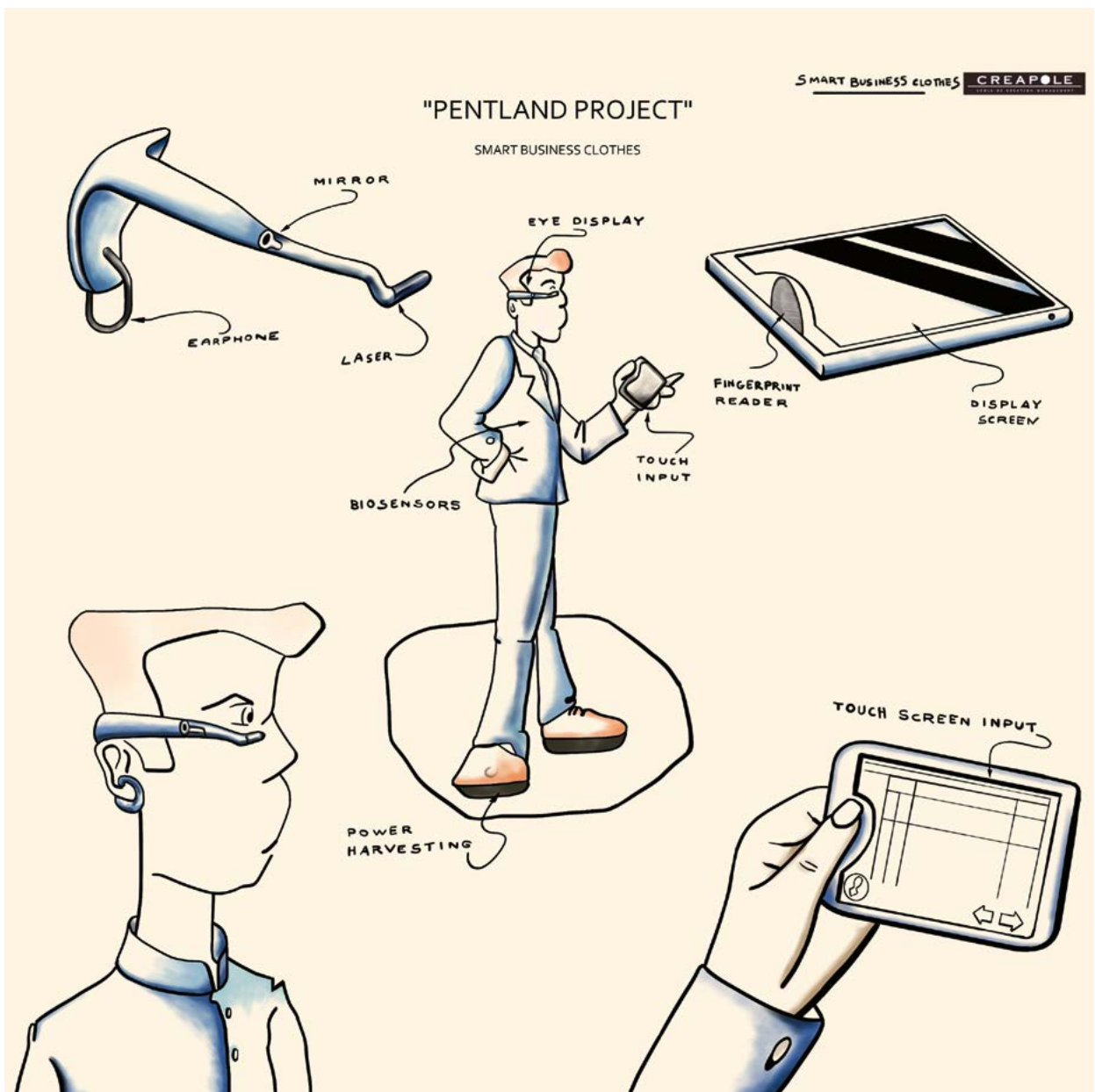
Perhaps the biggest mistake made by Western society is our adherence to the idea of ourselves as "rational individuals."

The foundations of modern Western society, and of rich societies everywhere, were laid in the 1700s in Scotland by Adam Smith, John Locke, and others. The understanding of ourselves that this early social science created is that humans are rational individuals, with "liberty and justice for all."<sup>3</sup> This is built in to every part of our society now—we use markets, we take direct democracy as our ideal of government, and our schools have dropped rhetoric classes to focus on training students to have better analytics skills.

But this rational individual model is wrong—and not just the rational part but, more importantly, the individual part. Our behavior is strongly influenced by those around us and, as we will see, the source of most of our wisdom. Our ability to thrive is due to learning from other people's experiences. We are not individuals but rather members of a social species. In fact, the idea of "rational individuals" reached its current form when mathematicians in the 1800s tried to make sense of Adam Smith's observation that people "...are led by an invisible hand to ... advance the interest of the society, and afford means to the multiplication of the species."<sup>4</sup> These mathematicians found that they could make the invisible hand work if they



Wearable electronics design, from collaboration between the author and Creapole Design School in Paris





used a very simplified model of human nature: people act only to benefit themselves (they are “rational”), and they act alone, independent of others (they are “individuals”).

What the mathematics of most economics and of most governing systems assume is that people make up their minds independently and they do not influence each other. That is simply wrong. While it may not be a bad first approximation, it fails in the end because it is people influencing each other, peer-to-peer, that causes financial bubbles, and cultural change, and (as we will see) it is this peer-to-peer influence that is the source of innovation and growth.

Furthermore, the idea of “rational individuals” is not what Adam Smith said created the invisible hand. Instead, Adam Smith thought: “It is human nature to exchange not only goods but also ideas, assistance, and favors ... it is these exchanges that guide men to create solutions for the good of the community.”<sup>5</sup> Interestingly, Karl Marx said something similar, namely that society is the sum of all of our social relationships.

The norms of society, the solutions for society, come from peer-to-peer communication—not from markets, not from individuals. We should focus on interaction between individuals, not individual genius. Until recently, though, we did not have the mathematics to understand and model such networks of peer-to-peer interaction. Nor did we have the data to prove how it all really works. Now we have both the maths and the data.

## **What the mathematics of most economics and most governing systems assume is that people make up their minds independently. That is simply wrong, because it is people influencing each other that causes financial bubbles, and it is this peer-to-peer influence that is the source of innovation and growth**

And so we come to the most fundamental question about ourselves: are we really rational individuals or are we more creatures of our social networks?

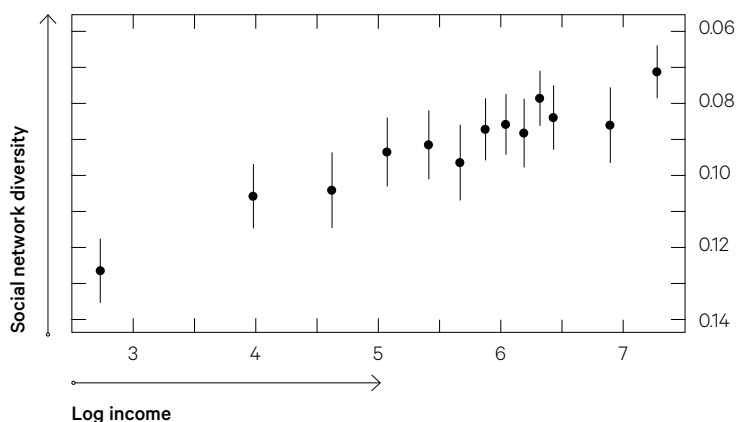
**Wealth Comes from Finding New Opportunities** To answer the question of who we really are, we now have data available from huge numbers of people in most every society on earth and can approach the question using very powerful statistical techniques. For instance, let us look at a sample of 100,000 randomly selected people in a mid-income country and compare their ability to hear about new opportunities (measured by how closed or open their social networks are) to their income.

The answer is that people who have more open networks make more money. Moreover, this is not just an artifact of the way we measured their access to opportunities, because you can get the same result looking at the diversity of jobs of the people they interact with, or the diversity of locations of the people that they interact with. Shockingly, if you compare people who have a sixth-grade education to illiterate people, this curve moves only a little to the left. If you look at people with college educations, the curve moves only a little bit to the right. The variation that has to do with education is trivial when compared with the variation that has to do with diversity of interaction.

You may wonder if greater network diversity *causes* greater income or whether it is the other way around. The answer is yes: greater network diversity causes greater income on



average (this is the idea of weak ties bringing new opportunities) but it is also true that greater income causes social networks to be more diverse. This is not the standard picture that we have in our heads when we design society and policy.



As people interact with more diverse communities, their income increases (100,000 randomly chosen people in mid-income country) (Jahani et al., 2017)

In Western society, we generally assume that individual features far outweigh social network factors. While this assumption is incorrect, nevertheless it influences our approach to many things. Consider how we design our schools and universities. My research group has worked with universities in several different countries and measured their patterns of social interactions. We find that social connections are dramatically better predictors of student outcome than personality, study patterns, previous training, or grades and other individual traits. Performance in school, in school tests, has more to do with the community of interactions that you have than with the things that the “rational individual” model leads us to assume are important. It is shocking.

It is better to conceive of humans as a species who are on a continual search for new opportunities, for new ideas, and their social networks serve as a major, and perhaps the greatest, resource for finding opportunities. The bottom line is that humans are like every other social species. Our lives consist of a balance between the habits that allow us to make a living by exploiting our environment and exploration to find new opportunities.

In the animal literature this is known as *foraging behavior*. For instance, if you watch rabbits, they will come out of their holes, they will go get some berries, they will come back every day at the same time, except some days they will scout around for other berry bushes. It is the tension between exploring, in case your berry bush goes away, and eating the berries while they are there.

This is exactly the character of normal human life. When we examined data for 100 million people in the US, we saw that people are immensely predictable. If I know what you do in the morning, I can know, with a ninety-percent-plus odds of being right, what you will

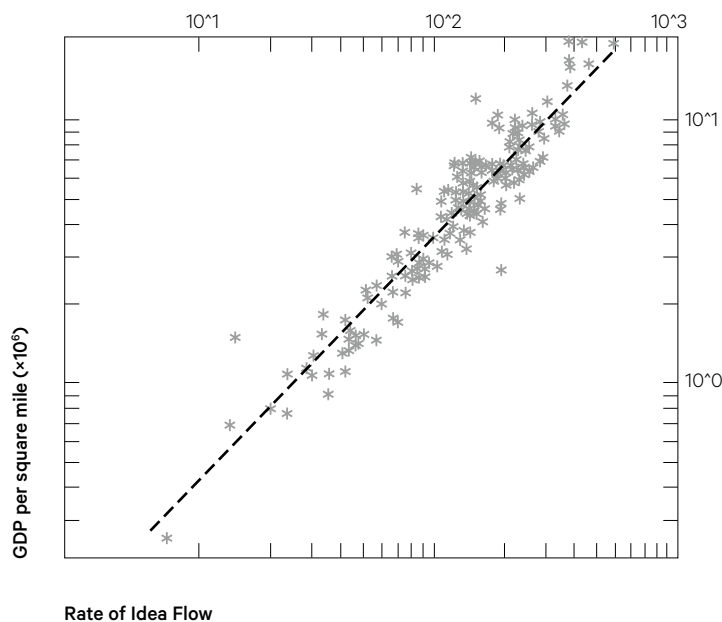
be doing in the evening and with whom. But, every once in a while, people break loose and they explore people and places that they visit only very occasionally, and this behavior is extremely unpredictable.

Moreover, when you find individuals who do not show this pattern, they are almost always sick or stressed in some way. You can tell whether a person's life is healthy in a general sense—both mental and physical—by whether they show this most basic biological rhythm or not. In fact, this tendency is regular enough that one of the largest health services in the US is using this to keep track of at-risk patients.



## Humans, as a species, are on a continual search for new opportunities, for new ideas, and their social networks serve as a major, and perhaps the greatest, resource for finding opportunities

If you combine this idea of foraging for novelty with the idea that diverse networks bring greater opportunities and greater income, you would expect that cities that facilitate connecting with a wide range of people would be wealthier. So, we gathered data from 150 cities in the US and 150 cities in the EU and examined the patterns of physical interactions between people.



As face-to-face communication within a city allows interaction between more diverse communities, the city wealth increases. Data from 150 cities in the US and 150 in the EU (Pan et al, 2013)

\* data    — — — our model



If your city facilitates more diverse interactions, then you likely have more diverse opportunities and over the long term you will make more money. From the figure above, you can see this model predicts GDP per square kilometer extremely accurately, in both the US and the EU. What that says is that the factors that we usually think about—investment, education, infrastructure, institutions—may be epiphenomenal. Instead of being the core of growth and innovation, they may make a difference primarily because they help or hinder the search for new opportunities. The main driver of progress in society may be the search for new opportunities and the search for new ideas—as opposed to skills in people’s heads or capital investment.

Summary: this new computational social science understanding of human behavior and society in terms of networks, the search for new opportunities, and the exchange of ideas might best be called Social Physics, a name coined two centuries ago by Auguste Comte, the creator of sociology. His concept was that certain ideas shaped the development of society in a regular manner. While his theories were in many ways too simplistic, the recent successes of computational social science show that he was going in the right direction. It is the *flow* of ideas and opportunities between people that drives society, providing quantitative results at scales ranging from small groups, to companies, cities, and even entire countries.

**Optimizing Opportunity** Once we understand ourselves better, we can build better societies. If the search for new opportunities and ideas is the core of human progress, then we should ask how best to accomplish this search. To optimize opportunity, I will turn to the science of financial investment, which provides clear, simple and well-developed examples of the trade-off between exploiting known opportunities and exploring for new ones.

In particular I will look at Bayesian portfolio analysis. These methods are used to choose among alternative actions when the potential for profit is unknown or uncertain (Thompson, 1933). Many of the best hedge funds and the best consumer retail organizations use this type of method for their overall management structure.

The core idea associated with these analysis methods is that when decision-makers are faced with a wide range of alternative actions, each with unknown payoff, they have to select actions to discover those that lead to the best payoffs, and at the same time exploit the actions that are currently believed to be the best in order to remain competitive against opponents. This is the same idea as animals foraging for food, or people searching for new opportunities while still making a living.

## **To optimize opportunity, I will turn to the science of financial investment, which provides clear, simple and well-developed examples of the trade-off between exploiting known opportunities and exploring for new ones**

In a social setting the payoff for each potential action can be easily and cheaply determined by observing the payoffs of other members of a decision-maker’s social network. This use of social learning dramatically improves both overall performance and reduces the cognitive load placed on the human participants. The ability to rapidly communicate and observe other decisions across the social network is one of the key aspects of optimal social learning and exploration for opportunity.



As an example, my research group recently examined how top performers maximize the sharing of strategic information within a social network stock-trading site where people can see the strategies that other people choose, discuss them, and copy them. The team analyzed some 5.8 million transactions and found that the groups of traders who fared the best all followed a version of this social learning strategy called “distributed Thompson sampling.” It was calculated that the forecasts from groups that followed the distributed Thompson sampling formula reliably beat the best individual forecasts by a margin of almost thirty percent. Furthermore, when the results of these groups were compared to results obtained using standard artificial intelligence (AI) techniques, the humans that followed the distributed Thompson sampling methodology reliably beat the standard AI techniques!

## **The use of social learning dramatically improves both overall performance and reduces the cognitive load placed on the human participants. The ability to rapidly communicate and observe other decisions across the social network is one of the key aspects of optimal social learning and exploration for opportunity**

It is important to emphasize that this approach is qualitatively the same as that used by Amazon to configure its portfolio of products as well as its delivery services. A very similar approach is taken by the best financial hedge funds. Fully dynamic and interleaved planning, intelligence gathering, evaluation, and action selection produce a powerfully optimized organization.

This social learning approach has one more advantage that is absolutely unique and essential for social species: the actions of the individual human are both in their best interest *and* in the best interest of everyone in the social network. Furthermore, the alignment of individuals’ incentives and the organization’s incentives are visible and understandable. This means that it is easy, in terms of both incentives and cognitive load, for individuals to act in the best interests of the society: optimal personal and societal payoffs are the same, and individuals can learn optimal behavior just by observing others.

### **Social Bridges in Cities: Paths to Opportunity**

Cities are a great example of how the process of foraging for new opportunities shapes human society. Cities are major production centers of society, and, as we have already seen, cities where it is easy to search for new opportunities are wealthier. Long-term economic growth is primarily driven by innovation in the society, and cities facilitate human interaction and idea exchange needed for good ideas and new opportunities.

These new opportunities and new ideas range from changes in means of production or product types to the most up-to-the-minute news. For example, success on Wall Street often involves knowing new events minutes before anyone else. In this environment, the informational advantages of extreme spatial proximity become very high. This may explain why Wall Street remains in a tiny physical area in the tip of Manhattan. The spatial concentration of



economic actors increases productivity at the firm level by increasing the flow of new ideas, both within and across firms.

Our evidence suggests that bringing together people from diverse communities will be the best way to construct a vibrant, wealthy city. When we examine flows of people in real cities, however, we find that mixing is much more limited than we normally think. People who live in one neighborhood work mostly with people from only a couple of other neighborhoods and they shop in a similarly limited number of areas.

Physical interaction is mostly limited to a relatively small number of *social bridges* between neighborhoods. It is perhaps unsurprising that people who spend time together, whether at work or at play, learn from each other and adopt very similar behaviors. When I go to work, or to a store, I may see someone wearing a new style of shoe, and think “Hey, that looks really cool. Maybe I’ll get some shoes like that.” Or, perhaps I go to the restaurant that I always like, and somebody nearby orders something different, and I think “Oh, that looks pretty good. Maybe I’m going to try that next time.” When people spend time together they begin to mimic each other, they learn from each other, and they adopt similar behaviors and attitudes.

In fact, what you find is that all sorts of behaviors, such as what sort of clothes they buy, how they deal with credit cards, even diseases of behavior (like diabetes or alcoholism), flow mostly within groups connected by social bridges. They do not follow demographic boundaries nearly as much. In a recent study of a large European city, for instance, we found that social bridges were three hundred percent better at predicting people’s behaviors than demographics, including age, gender, income, and education

Groups of neighborhoods joined by rich social bridges form local cultures. Consequently, by knowing a few places that a person hangs out in, you can tell a huge amount about them. This is very important in both marketing and politics. The process of learning from each other by spending time together means that ideas and behaviors tend to spread mostly within the cluster, but not further. A new type of shoe, a new type of music, a political viewpoint, will spread within a cluster of neighborhoods joined by social bridges, but will tend not to go across cluster boundaries to other places. Advertisers and political hacks talk about influencers changing people’s minds. What I believe is that it is more about social bridges, people hanging out, seeing each other, interacting with each other, that determines the way ideas spread.

## **Long-term economic growth is primarily driven by innovation, and cities facilitate human interaction and idea exchange needed for good ideas and new opportunities**

Summary: what Adam Smith said about people exchanging ideas—that it is the peer-to-peer exchanges that determine norms and behaviors—is exactly true. But what is truly stunning is that because we hold onto the “rational individual” model of human nature, we assume that preferences are best described by individual demographics—age, income, gender, race, education, and so forth—and that is wrong. The way to think about society is in terms of these behavior groups. Who do they associate with? What are those other people doing? The idea of social bridges is a far more powerful concept than demographics, because social bridges are the most powerful way that people influence each other. By understanding the social bridges in society, we can begin to build a smarter, more harmonious society.



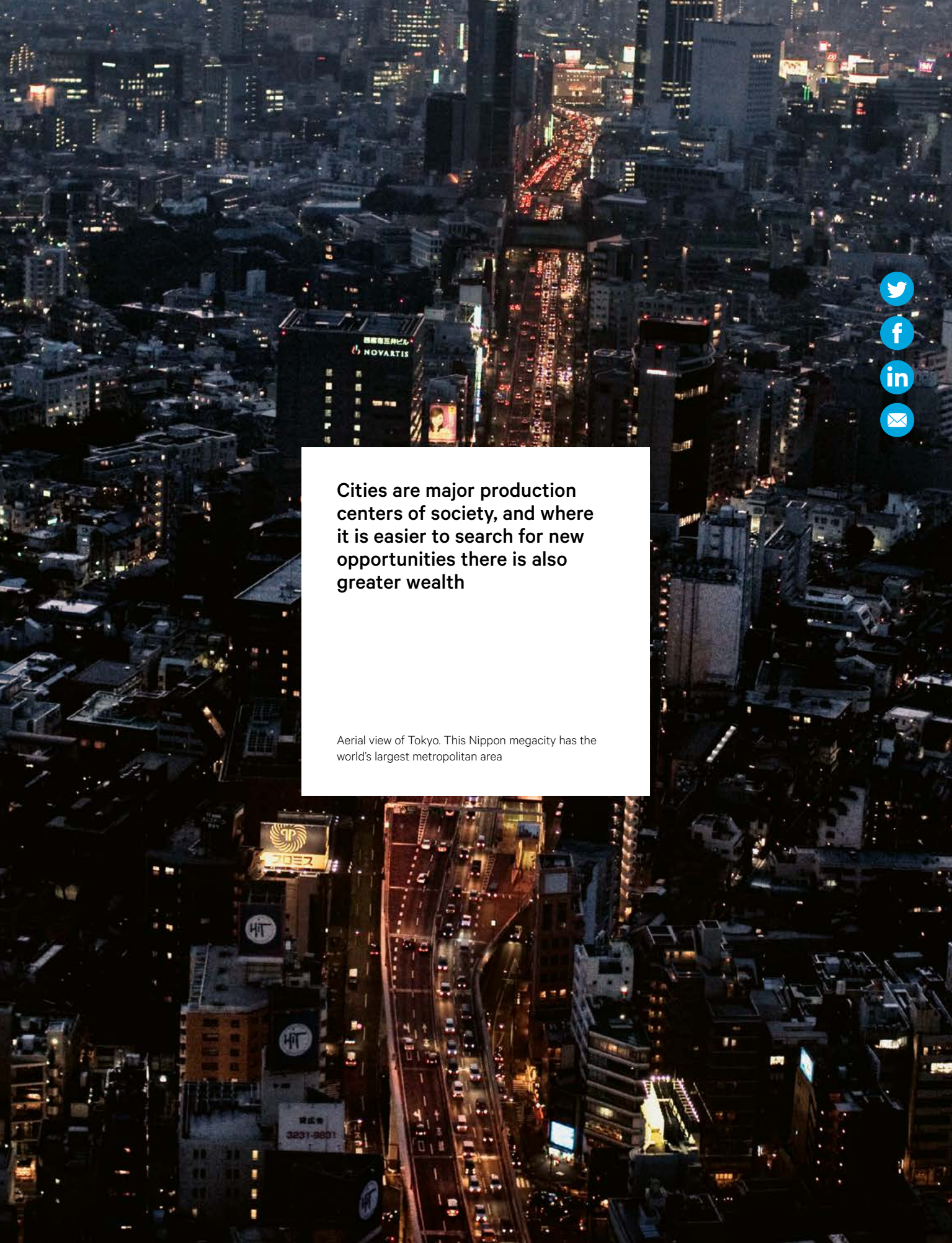


**Social bridges are three-hundred percent better at predicting people's behaviors than demographics**

*Influencer Susie Bubble (r.) before the Louis Vuitton fashion show at the Paris Fashion Week Womenswear Spring/Summer 2019*







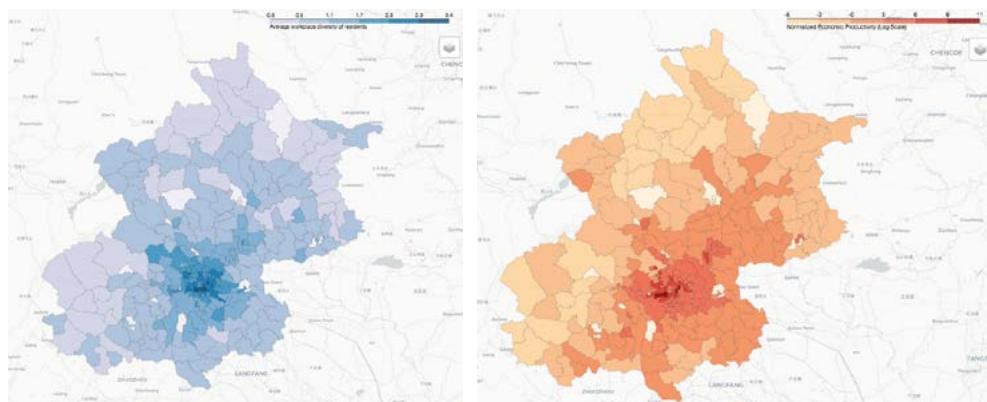
Cities are major production centers of society, and where it is easier to search for new opportunities there is also greater wealth

Aerial view of Tokyo. This Nippon megacity has the world's largest metropolitan area



**Diversity and Productivity** While there has been a sharp increase in remote, digital communications in modern times, physical interactions between people remain the key medium of information exchange. These social interactions between people include social learning through observation (for example, what clothes to wear or what food to order) and through intermediary interactions (for example, word of mouth, mutual friends). To infer interactions between people, a network of interaction was first obtained based on the individual's proximity. Physical proximity has been shown to increase the likelihood of face-to-face conversations, and this increase in interaction strength is widely used to help city planning and placement of city amenities.

The social bridges idea is that individuals can commonly be identified as part of a community based on where they live, work, and shop. Where you invest your most valuable resource—time—reveals your preferences. Each community typically has access to different pools of information, opportunities, or offers different perspectives. Diverse interactions should then increase a population's access to the opportunities and ideas required for productive activity.



Interaction diversity (left) and future economic growth (right) for the city of Beijing; it can be seen that they are highly correlated (Chong, Bahrami, and Pentland, 2018)

When we apply this logic to the cities in the US, Asia, and Europe, we find that the effect of this sort of interaction diversity has an effect comparable to that of increasing population. In other words, it is not just the number of individuals in the region that predicts economic growth, but also idea flow via the social bridges that connect them. If we compare the explanatory strength of interaction diversity with other variables such as average age or percentage of residents who received a tertiary education, we find that these traditional measures are much weaker at explaining economic growth than social bridge diversity. This means that models and civil systems that depend only on factors such as population and education may be missing the main effects.



## A New Social Contract



In 2014 a group of big data scientists (including myself), representatives of big data companies, and the heads of National Statistical Offices from nations in both the north and south met within the United Nations headquarters and plotted a revolution. We proposed that all the nations of the world measure poverty, inequality, injustice, and sustainability in a scientific, transparent, accountable, and comparable manner. Surprisingly, this proposal was approved by the UN General Assembly in 2015, as part of the 2030 Sustainable Development Goals.

This apparently innocuous agreement is known as the Data Revolution within the UN, because for the first time there is an international commitment to discover and tell the truth about the state of the human family as a whole. Since the beginning of time, most people have been isolated, invisible to government and without information about or input to government health, justice, education, or development policies. But in the last decade this has changed. As our UN Data Revolution report, titled “A World That Counts” put it:

Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time, designing, monitoring and evaluating effective policies becomes almost impossible. New technologies are leading to an exponential increase in the volume and types of data available, creating unprecedented possibilities for informing and transforming society and protecting the environment. Governments, companies, researchers and citizen groups are in a ferment of experimentation, innovation and adaptation to the new world of data, a world in which data are bigger, faster and more detailed than ever before. This is the data revolution.<sup>6</sup>

More concretely, the vast majority of humanity now has a two-way digital connection that can send voice, text, and, most recently, images and digital sensor data because cellphone networks have spread nearly everywhere. Information is suddenly something that is potentially available to everyone. The Data Revolution combines this enormous new stream of data about human life and behavior with traditional data sources, enabling a new science of “social physics” that can let us detect and monitor changes in the human condition, and to provide precise, non-traditional interventions to aid human development.

Why would anyone believe that anything will actually come from a UN General Assembly promise that the National Statistical Offices of the member nations will measure human development openly, uniformly, and scientifically? It is not because anyone hopes that the UN will manage or fund the measurement process. Instead, we believe that uniform, scientific measurement of human development will happen because international development donors are finally demanding scientifically sound data to guide aid dollars and trade relationships.

Moreover, once reliable data about development starts becoming familiar to business people, we can expect that supply chains and private investment will start paying attention. A nation with poor measures of justice or inequality normally also has higher levels of corruption, and a nation with a poor record in poverty or sustainability normally also has a poor record of economic stability. As a consequence, nations with low measures of development are less attractive to business than nations with similar costs but better human development numbers.

## Building a Data Revolution

How are we going to enable this data revolution and bring transparency and accountability to governments worldwide?



The key is safe, reliable, uniform data about the human condition. To this end, we have been able to carry out country-scale experiments that have demonstrated that this is a practical goal. For instance, the Data for Development (D4D) experiments that I helped organize for Cote d'Ivoire and Senegal, each of which had the participation of hundreds of research groups from around the world, have shown that satellite data, cellphone data, financial transaction data, and human mobility data can be used to measure the sustainable development goals reliably and cheaply. These new data sources will not replace existing survey-based census data, but rather will allow this rather expensive data to be quickly extended in breadth, granularity, and frequency.<sup>7</sup>

But what about privacy? And won't this place too much power in too few hands? To address these concerns, I proposed the "New Deal on Data" in 2007, putting citizens in control of data that are about them and also creating a data commons to improve both government and private industry (see [http://hd.media.mit.edu/wef\\_globalit.pdf](http://hd.media.mit.edu/wef_globalit.pdf)). This led to my co-leading a World Economic Forum discussion group that was able to productively explore the risks, rewards, and cures for these big data problems. The research and experiments I led in support of this discussion shaped both the US Consumer Privacy Bill of Rights, the EU Data Protection laws, and is now helping China determine its data protection policies. While privacy and concentration of power will always be a concern, as we will shortly see there are good solutions available through a combination of technology standards (for example, "Open Algorithm" described below) and policy (e.g., open data including aggregated, low-granularity data from corporations).

## **D4D experiments have shown that satellite data, cellphone data, financial transaction data, and human mobility data can be used to measure sustainable development goals reliably and cheaply**

Following up on the promise uncovered by these academic experiments and World Economic Forum discussions, we are now carrying out country-scale pilots using our Open Algorithms (OPAL) data architecture in Senegal and Colombia with the help of Orange S.A., the World Bank, Paris21, the World Economic Forum, the Agence Française de Développement, and others. The Open Algorithms (OPAL) project, developed originally as an open source project by my research group at MIT (see <http://trust.mit.edu>), is now being deployed as a multi-partner socio-technological platform led by Data-Pop Alliance, Imperial College London, the MIT Media Lab, Orange S.A., and the World Economic Forum, that aims to open and leverage private sector data for public good purposes (see <http://opalproject.org>). The project came out of the recognition that accessing data held by private companies (for example, call detail records collected by telecom operators, credit card transaction data collected by banks, and so on) for research and policy purposes requires an approach that goes beyond ad hoc data challenges or through nondisclosure agreements. These types of engagements have offered ample evidence of the promise, but they do not scale nor address some of the critical challenges, such as privacy, accountability, and so on.

OPAL brings stakeholders together to determine what data—both private and public—should be made accessible and used for which purpose and by whom. Hence, OPAL adds a new dimension and instrument to the notion of "social contract," namely the view that moral and/or political obligations of people are dependent upon an agreement among them to form the society in which they live. Indeed, it provides a forum and mechanism for deciding what



Surveillance cameras at the Paul-Loebe-Hause, home to Budestag members and headquarters of the European Commission. Berlin, Germany, April 2018





levels of transparency and accountability are best for society as a whole. It also provides a natural mechanism for developing evidence-based policies and a continuous monitoring of the various dimensions of societal well-being, thus offering the possibility of building a much deeper and more effective science of public policy.

OPAL is currently being deployed through pilot projects in Senegal and Colombia, where it has been endorsed by and benefits from the support of their National Statistical Offices and major local telecom operators. Local engagement and empowerment will be central to the development of OPAL: needs, feedback, and priorities have been collected and identified through local workshops and discussions, and their results will feed into the design of future algorithms. These algorithms will be fully open, therefore subject to public scrutiny and redress. A local advisory board is being set up to provide guidance and oversight to the project. In addition, trainings and dialogs will be organized around the project to foster its use and diffusion as well as local capacities and awareness more broadly. Initiatives such as OPAL have the potential to enable more human-centric accountable and transparent data-driven decision-making and governance.

In my view, OPAL is a key contribution to the UN Data Revolution—it is developing and testing a practical system that insures that every person counts, their voices are heard, their needs are met, their potentials are realized, and their rights are respected.

**Smarter Civic Systems** Using the sort of data generated by OPAL, computational social science (CSS) researchers have shown that Jane Jacobs, the famous urban visionary and advocate of the 1900s, was right: cities are engines of innovation, powered by the interaction of diverse communities. The question, then, is how to best harness this innovation. Today we use markets and market forces to separate good ideas from bad, and to grow the good ones to scale.

But, as CSS has shown, markets are a bad way to characterize human society. They are based on greedy optimization and ignore the power of human social processes. The only reason that they work at all may be that regulators “tweak” them to suit the regulators’ preconceived notions of good and bad. Direct democracy suffers from the same problems, which is why most countries have representative democracy instead. Unfortunately, representatives in such systems are all too prone to capture by special interests.

Another path to harnessing human innovations is suggested by the insight that powers today’s cutting-edge artificial intelligence (AI) algorithms. Most of today’s AI starts with a large, random network of simple logic machines. The random network then learns by changing the connections between simple logic machines (“neurons”), with each training example slightly changing the connections between all the neurons based on how the connection contributes to the overall outcome. The “secret sauce” is the learning rule, called the *credit assignment function*, which determines how much each connection contributed to the overall answer. Once the contributions of each connection have been determined, then the learning algorithm that builds the AI is simple: connections which have contributed positively are strengthened; those that have contributed negatively are weakened.

This same insight can be used to create a better human society, and, in fact, such techniques are already widely used in industry and sports, as I will explain. To understand this idea, imagine a human organization as a kind of brain, with humans as the individual neurons. Static firms, symbolized by the ubiquitous organization chart, have fixed connections and, as a result, a limited ability to learn and adapt. Typically, their departments become siloed, with little communication between them so that the flow of fresh, crosscutting ideas is blocked. As a consequence, these statically configured, minimally interconnected organizations risk falling to newer, less ossified competitors.





But if an organization's skills can be supercharged by adopting the right sort of credit assignment function, then the connections, among individuals, teams, and teams of teams, might continuously reorganize themselves in response to shifting circumstances and challenges. Instead of people being forced to be simple rule-following machines, people would engage in continuous improvement that is seen in the *Kaizen-style* manufacturing developed by Toyota, the "quality team" feedback approach adopted by many corporations, or the continuous, data-driven changes in sports team rosters described in the book *Moneyball*.

## **Another path to harnessing human innovations is suggested by the insight that powers today's cutting-edge AI algorithms. The "secret sauce" is the learning rule, called the *credit assignment function*, which determines how much each connection contributed to the overall answer**

The key to such dynamic organizations is continuous, granular, and trustworthy data. You need to know what is actually happening *right now* in order to continuously adapt. Having quarterly reports, or conducting a national census every ten years, means that you cannot have an agile learning organization or a learning society. OPAL and the data revolution provide the foundation upon which a responsive, fair, and inclusive society can be built. Without such data we are condemned to a static and paralyzed society that will necessarily fail the challenges we face.

Today, online companies such as Amazon and Google, and financial services firms such as Blackrock and Renaissance, have enthusiastically adopted and developed these data-driven agile approaches to manage their organizations. Obviously, such continuously adapting and learning organizations, even if assisted by sophisticated algorithms, will require changing policy, training, doctrine, and a wide spectrum of other issues.

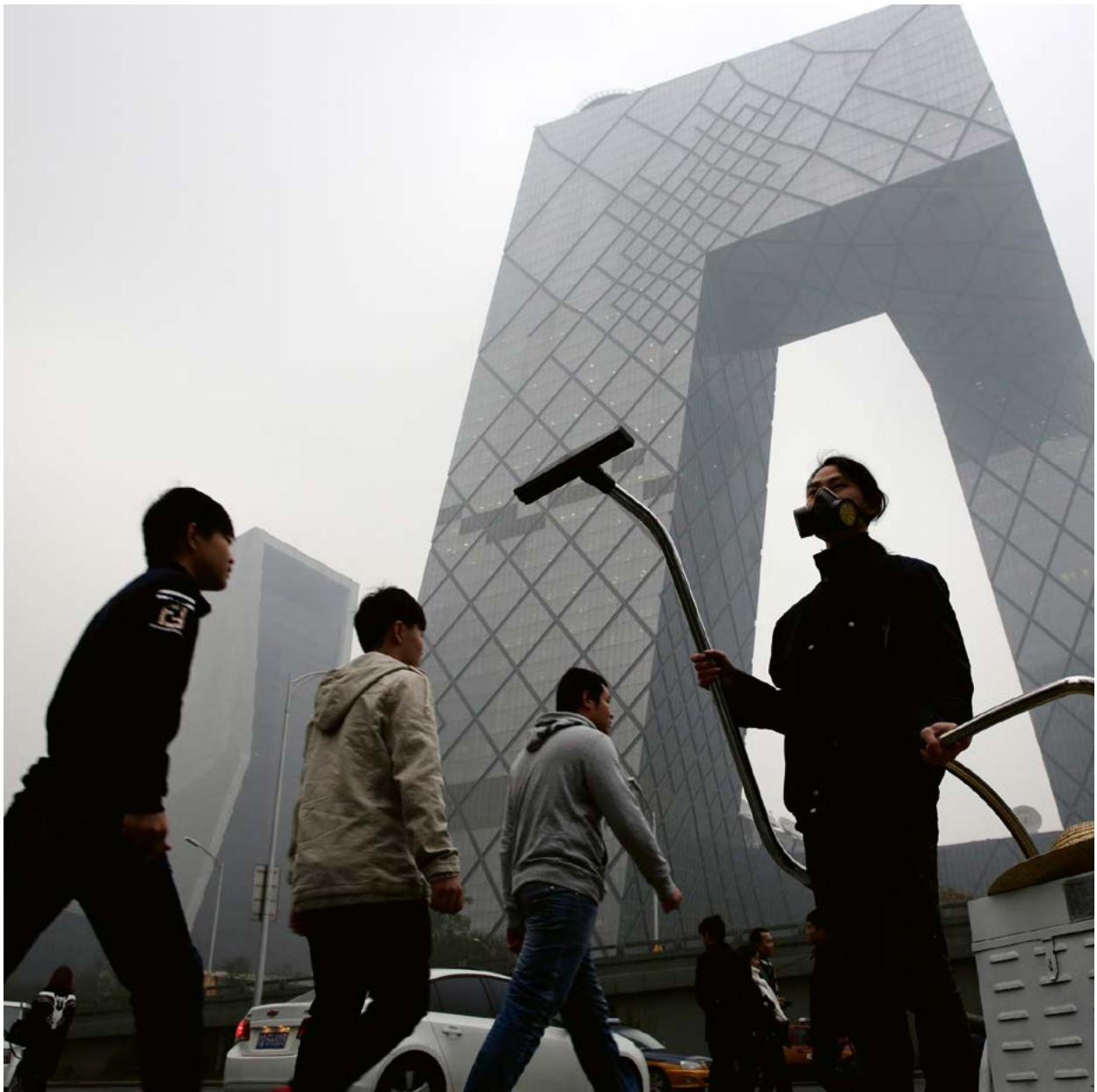
**Avoiding Chaos** To most of us change means chaos. So how can we expect to achieve this vision of a dynamic society without continuous chaos? The key is to have a credit assignment function that both makes sense for each individual and yet at the same time yields global optimal performance. This would be a *real* version of the "invisible hand," one that works for real people and not just for rational individuals.

The phrase *makes sense* means that each individual must be able to easily understand their options and how to make choices that are good for them as an individual, and also understand how it is that what is good for them is also good for the whole organization. They must be able to easily and clearly communicate the choices they have made and the anticipated or actual outcomes resulting from those choices. Only through this sort of understanding and alignment of incentives within society will the human participants come to both trust and helpfully respond to change.

Sound impossible? It turns out that people already quite naturally do this in day-to-day life. As described in the first section of this paper, people rely on social learning to make decisions. By correctly combining the experiences of others we can both adapt to rapidly changing circumstances and make dramatically better day-to-day decisions. Earlier I described how people naturally did this in a social financial platform.



A young man collects smog in Beijing with an industrial vacuum cleaner in order to make "smog bricks" that will be recycled as building materials. With this 100-day project, he sought to demonstrate the health and environmental effects of urban pollution. Beijing, November 2015





The keys to successful rapid change are granular data about what other people are doing and how it is working out for them, and the realization that we cannot actually act independently because every action we take affects others and feeds back to ourselves as others adapt to our actions. The enemies of success are activities that distort our perception of how other people live their lives and how their behaviors contribute to or hinder their success, for example, advertising, biased political advocacy, segregation, and the lack of strong reputational mechanisms all interfere with rapid and accurate social learning.

**The Human Experience** How can we begin to encourage such learning organization and societies? The first steps should be focused on accurate, real-time knowledge about the successes and failures that others have experienced from implementing their chosen strategies and tactics. In prehistoric times people knew quite accurately what other members of their village did and how well it worked, and this allowed them to quickly develop compatible norms of behavior. Today we need digital mechanisms to help us know about what works and what does not work. This sort of reputation mechanism is exactly the benefit of using an architecture like OPAL.

Strong reputation mechanisms allow local community determination of norms and regulations instead of laws created and enforced by elites. For instance, frontline workers often have better ideas about how to deal with challenging situations than managers, and tactical engineers know more about how a new capability is shaping up than its designers do. The secret to creating an agile, robust culture is closing the communications gap between people who do and people who organize, so that employees are both helping to create plans and executing them. This closure fits with another key finding: developing the best strategy in any scenario involves striking a balance between engaging with familiar practices and exploring fresh ideas.

Actively encouraging sharing in order to build community knowledge offers another benefit: when people participate and share ideas, they feel more positive about belonging to the community and develop greater trust in others. These feelings are essential for building social resilience. Social psychology has documented the incredible power of group identities to bond people and shape their behavior; group membership provides the social capital needed to see team members through inevitable conflicts and difficult periods.

### Summary: A New Enlightenment

Many of the traditional ideas we have about ourselves and how society works are wrong. It is not simply the brightest who have the best ideas; it is those who are best at harvesting ideas from others. It is not only the most determined who drive change; it is those who most fully engage with like-minded people. And it is not wealth or prestige that best motivates people; it is respect and help from peers.

The disconnect between traditional ideas about our society and the current reality has grown into a yawning chasm because of the effects of digital social networks and similar technology. To understand our new, hyper-connected world we must extend familiar economic and political ideas to include the effects of these millions of digital citizens learning from each other and influencing each other's opinions. We can no longer think of ourselves as only rational individuals reaching carefully considered decisions; we must include the dynamic social networking effects that influence our individual decisions and drive economic bubbles, political revolutions, and the Internet economy. Key to this are strong and rapid reputation mechanisms, and inclusiveness in both planning and execution.

Today it is hard to even imagine a world where we have reliable, up-to-the-minute data about how government policies are working and about problems as they begin to develop. Perhaps the most promising uses for big data are in systems like OPAL, which allow statisticians in government statistics departments around the world to make a more accurate, “real-time” census and more timely and accurate social surveys. Better public data can allow both the government and private sectors to function better, and with these sorts of improvements in transparency and accountability we can hope to build a world in which our government and social institutions work correctly.

Historically we have always been blind to the living conditions of the rest of humanity; violence or disease could spread to pandemic proportions before the news would make it to the ears of central authorities. We are now beginning to be able to see the condition of all of humanity with unprecedented clarity. Never again should it be possible to say: “We didn’t know.”





## Notes

1. "Secret signals," Mark Buchanan, *Nature* 457, 528–530, 2009.
2. "Life in the Network: the coming age of computational social science," David Lazer, Alex Pentland, et al., *Science*. 2009 Feb 6; 323(5915): 721–723.
3. From the USA's "Pledge of Allegiance."
4. Adam Smith, *The Theory of Moral Sentiments*, 1759.
5. Ibid.
6. At <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>.
7. It is worth mentioning that this buries the classic arguments against utilitarianism: the nations of the world have agreed that you can in fact measure the quality of human life.

## Select Bibliography

- Chong, Shi, Bahrami, Mohsen, and Pentland, Alex. 2018. *A Computational Approach to Urban Economics*. In preparation.
- Jahani, Eaman, Sundsøy, Pål, Bjelland, Johannes, Bengtsson, Linus, Pentland, Alex "Sandy," and de Montjoye, Yves-Alexandre. 2017. "Improving official statistics in emerging markets using machine learning and mobile phone data." *EPJ Data Science* 6:3. <https://doi.org/10.1140/epjds/s13688-017-0099-3> (accessed on September 19, 2018).
- Pan, Wei, Ghoshal, Gourab, Krumme, Coco, Cebrian, Manuel, and Pentland, Alex. 2013. "Urban characteristics attributable to density-driven tie formation." *Nature Communications* 4 article number: 1961.
- Thompson, William R. 1933. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples." *Biometrika* 25(3–4): 285–294. <https://doi.org/10.1093/biomet/25.3-4.285> (accessed on September 19, 2018).



**Sandip Tiwari**  
Cornell University

Sandip Tiwari is Charles N. Mellowes Professor in Engineering at Cornell University with interests in the engineering and science underlying information processing. His current explorations are in understanding and laying foundations of the broad themes that make machine learning a useful approximate proxy for understanding the natural world's dynamics. This includes using probabilistic and neural network techniques to understand complexity, energetics, entropies, and confidence in inferences as theoretical tools, and nanoscale hardware approaches as practical tools. Recognitions of his past contributions include the Clelio Brunetti Award and Fellowship of the Institute of Electrical and Electronics Engineers, the Young Scientist Award from the Institute of Physics, and Fellowship of the American Physical Society. One of his coauthored patents underlies a common molecular charge profiling technique employed in commercial gene sequencing tools. He believes passionately in education and research as the apex of human pursuits, bringing together curiosity and collective efforts for community well-being when combined with open sharing, critical self-reflection, and community debates of the technical pursuits.

Recommended book: *No Small Matter: Science on the Nanoscale*, Felice C. Frankel and George M. Whitesides, Belknap Press, 2009.

**Nanotechnology's influence in our daily life is reflected in mobile communication tools, medical diagnostics and new treatments, the use of data by companies and governments, and its accumulation in the cloud. Society reacts slowly to rapidly unfolding technological changes. Nanotechnology, with its atomic-scale capabilities in which much of the natural and physical world's dynamics unfold, has the potential to make far more dramatic leaps than humanity has encountered in the past. Evolutionary changes—the natural world's manipulation, such as through genes—and emergent changes—the physical world's manipulation and autonomy, such as through artificial intelligence—can now be brought together to cause profound existential changes. The complex existential machines thus created have a ghost in them, and humanity needs to shape technology at each stage of the machine building so that it is an angelic ghost.**

Even some of your materialistic countrymen are prepared to accept—at least as a working hypothesis—that some entity has—well, invaded Hal. Sasha has dug up a good phrase: “The Ghost in the Machine.”

Arthur C. Clarke (in *2010: Odyssey Two*, 1982)

The official doctrine, which hails chiefly from Descartes, is something like this. ... every human being has both a body and a mind. Some would prefer to say that every human being is both a body and a mind. His body and his mind are ordinarily harnessed together, but after the death of the body his mind may continue to exist and function. ... Such in outline is the official theory. I shall often speak of it, with deliberate abusiveness, as ‘the dogma of the Ghost in the Machine.’ I hope to prove that it is entirely false, and false not in detail but in principle.

Gilbert Ryle (in *The Concept of Mind*, 1949)



Six years ago, in an earlier article<sup>1</sup> in this book series, we explored the implications of the ability to observe and exercise control at the atomic and molecular scale, also known as nanotechnology. We had concluded with speculations and numerous questions regarding the physical and biological complexity that the future will unfold as a result of this technology’s development. Our reasoning was that nature’s complexity arose in the interactions that take place at the atomic scale—atoms building molecules, complex molecules in turn leading to factories such as from the cells for reproduction, and the creation of multifaceted hierarchical systems. Physical laws, such as the second law of thermodynamics, still hold good, so this development of complexity takes place over long time scales in highly energy-efficient systems. The use of nanotechnology’s atomic-scale control—nature’s scale—has brought continuing improvements in efficiency in physical systems and a broadening of its utilization in biology and elsewhere. This ability to take over nature’s control by physical technology gives humans the ability to intervene beneficially, but also to raise existential questions about man and machine. Illustrative examples were *emergent machines* as self-replicating automatons where hardware and software have fused as in living systems, or *evolution machines* where optimization practiced by engineering modifies the evolutionary construct. Computational machines now exist as the emergent variety which improve themselves as they observe and manipulate more data, retune themselves by changing how the hardware is utilized, and span copies of themselves through partitioning over existent hardware, even if they are not yet building themselves physically except in rudimentary 3D printing. Numerous chemicals and drugs are now built via cells and enzymes as the biological factories.

The existential questions that we concluded with in the article have only buttressed themselves in this interregnum. Rapid development of CRISPR (clustered regularly interspaced short palindromic repeats) and of machine learning evolving to artificial intelligence (AI)<sup>2</sup> have brought us to a delicate point.

This paper—scientific and philosophical musing—underscores lessons from these intervening years to emphasize the rapid progress made, and then turns to what this portends. We look back at how the physical principles have influenced nanotechnology’s progress and evolution, where it has rapidly succeeded and where not, and for what reasons. As new technologies appear and cause rapid change, society’s institutions and us individually are slow in responding toward accentuating the positive and suppressing the negatives through the restraints—community-based and personal—that bring a healthy balance. Steam engines, when invented, were prone to explosions. Trains had accidents due to absent signaling sys-

tems. Society put together mechanisms for safety and security. Engines still explode, accidents still happen, but at a level that society has deemed acceptable. We are still working on the control of plastics, and failing at global warming. These latter reflect the long latency of community and personal response.

In the spirit of the past essay, this paper is a reflection on the evolution of the nanotechnology catalyzed capabilities that appear to be around the corner, as well as the profound questions that society needs to start reflecting and acting on.



## **The use of nanotechnology's atomic-scale control—nature's scale—has brought continuing improvements in efficiency in physical systems and a broadening of its utilization in biology and elsewhere. This ability to take over nature's control by physical technology gives humans the ability to intervene beneficially, but also to raise existential questions about man and machine**

This past decadal period has brought numerous passive applications of nanotechnology into general use. As a material drawing on the variety of unique properties that can be achieved in specific materials through small-scale phenomena, usage of nanotechnology is increasingly pervasive in numerous products, though still at a relatively high fiscal cost. These applications range from those that are relatively simple to others that are quite complex. Coatings give strength and high corrosion and wear resistance. Medical implants—stents, valves, pacemakers, others—employ such coatings, as do surfaces that require increased resistance in adverse environments: from machine tools to large-area metal surfaces. Since a material's small size changes optical properties, trivial but widespread uses, such as sunscreens, or much more sophisticated uses, such as optically mediated interventions in living and physical systems, have become possible. The mechanical enhancements deriving from nanotubes have become part of the arsenal of lightweight composites. The nanoscale size and surface's usage has also made improved filtration possible for viruses and bacteria. Batteries employ the porosity and surface properties for improvements in utilitarian characteristics—energy density, charge retention, and so on—and the proliferation of electric cars and battery-based storage promises large-scale utilization. Surfaces make it possible to take advantage of molecular-scale interaction for binding and interaction. This same mechanism also allows targeting with specificity in the body for both enhanced detection as well as drug delivery. So, observation, in particular, of cancerous growths and their elimination has become an increasing part of the medical tool chest. This same surface-centric property makes sensors that can detect specific chemicals in the environment possible. The evolutionary machine theme's major role has consisted in its ability, through large-scale, nature-mimicking random trials conducted in a laboratory-on-a-chip, to sort, understand, and then develop a set of remedies that can be delivered to the target where biological mechanisms have gone awry. Cancer, the truly challenging frontier of medicine, has benefited tremendously through the observation and intervention provided by nanotechnology's tool chest even though this truly complex behavior (a set of many different diseases) is still very far from any solution except in a few of its forms. The evolutionary machine theme is also evident in the production methods used





Graphene slurry, containing graphene and polymer binders, at the National Graphene Institute facility, part of the University of Manchester, in Manchester, UK, April, 2018. Graphene is increasingly in demand for use in batteries



for a variety of simple-to-complex compounds—from the strong oxidant that is hydrogen peroxide via enzymes to the vaccine for Ebola virus produced via tobacco plants—that have become commonplace.

Admittedly, much of this nanotechnology usage has been in places where cost has been of secondary concern, and the specific property attribute of sufficient appeal to make a product attractive. This points to at least one barrier, that of the cost of manufacturing that has remained a challenge. Another has been the traditional problem of over-ebullience that drives market culture. Graphene, nanotubes, and other such forms are still in search of large-scale usage. An example of market culture is the space elevator based on carbon nanotubes that caught popular attention. Thermodynamics dictates the probabilities of errors—defects, for example—in assemblies. And although a nanotube, when small, can exhibit enormous strength, once one makes it long enough, even the existence of one defect has serious repercussions. So, space elevators remain science fiction.



## **The evolutionary machine theme's major role has consisted in its ability to sort, understand, and then develop a set of remedies that can be delivered to the target where biological mechanisms have gone awry**

There is one place, though, where the cost barrier continues to be overcome at quite an incredible rate. In human-centric applications, the evolution of electronics in the information industry (computation and communication) has seen the most dramatic cost reduction and expansion of widespread usage. The cellphone, now an Internet-accessing and video delivering smartphone, has been an incredibly beneficial transforming utility for the poor in the Third World.

The nature of archiving and accessing data has changed through the nanoscale memories and the nanoscale computational resources that increasingly exist in a far removed “non-place”—euphemistically in the cloud. These are accessed again by a nanoscale-enabled plethora of developments in communications, whether wireless or optical, and through the devices that we employ. Videos, texting, short bursts of communications, and rapid dissemination of information is ubiquitous. Appliances, large and small, are being connected together and controlled as an Internet of Things in a widening fabric at home and at work for desired goals such as reducing energy consumption, or improving health, or for taking routine tasks away from human beings.

Availability of information and the ability to draw inferences from it has brought machine learning, also known as artificial intelligence (AI), to the fore. Data can be analyzed and decisions made autonomously in a rudimentary form in the emergent machine sense. Artificial intelligence coupled to robotics—the physical use of this information—is also slowly moving from factories to human-centric applications, such as autonomous driving, in which sensing, inferencing, controlling, and operating all come together. This is all active usage that has an emergent and evolutionary character.

Quantum computation is another area that has made tremendous progress during this period. Quantum computing employs entangled superposition of information accessible at



the quantum scale, which becomes possible at the nanoscale, in order to proceed with the computation. It is a very different style than the traditional one of deterministic computing, in which bits are classical. Classical in the sense that they are either, say, “0” or “1” Boolean bits. We can transform them through computational functions, a “0” to “1”, for example by an inverter, or a collection of these—one number—and another collection of these—another number—through a desired function to another collection of these, which is another number. Adding or multiplying is such a functional operation. At each stage of computation, these “0”s and “1”s are being transformed deterministically. In classical computation, one cannot, in general, get back to the starting point once one has performed transformations since information is being discarded along the way. A multiplication product usually has multiple combinations of multiplicands and multipliers. Quantum bits as entangled superposed states are very different. A single quantum bit system is a superposition of “0” and “1”. We do not know which one it is except that it is one of them. When we measure it, we find out whether it is a “0” or a “1”. A two quantum bit system can be an entangled system where the two bits are interlinked. For example, it could be a superposition where if the first is a “0” the second is a “1”, and if the first is a “1” then the second is a “0”. This is an entangled superposition. Only when we make a measurement does one find out if it is the “01” combination or the “10” combination. A system composed of a large number of quantum bits can hold in it far more of these linkages, and one can manipulate these through the computation—without the measurement of the result—while the entanglement and its transformations continue to exist. While performing this computation, without having observed the result, one can actually go back to the starting point since no information has been discarded. So, the computation proceeds and only when we make a measurement, do we find out the result, which is in our classical world. It is this ability, and transformations in it while still keeping the entanglement and its related possibility options open—unlike the discarding of them in the classical mode—that endows quantum computing with properties that surpass those of classical computation. Fifty quantum bit systems exist now. This is almost the point in which quantum computation begins to achieve capabilities superior to that of classical computation. Numerous difficult problems, cryptography having been the earliest rationale for pursuing this approach, but also many more practical interesting problems—understanding molecules and molecular interactions, working toward further complexity of them including into drug discovery—become potentially solvable. Fifty quantum bits make much more complexity to be computed. Since nature is quantum in its behavior at the smallest scales where the diversity of interactions occurs, and classical behavior is a correspondence outcome, that is, a highly likely statistical outcome, quantum computing represents a way that we have now found to simulate how nature itself computes. A new style of computing is being born.

**Artificial intelligence coupled to robotics—the physical use of this information—is also slowly moving from factories to human-centric applications, such as autonomous driving, in which sensing, inferencing, controlling, and operating all come together**

These broad and pervasive changes—still in infancy—are very much on the same scale of changes that arose with the invention of the printing press and the invention of mechanized



transportation. The press democratized learning and information delivery. Mechanized transportation made the world smaller and eased the delivery of goods. Both were instrumental in making a more efficient way—in time, but also in energy, and in other dimensions—for human uplifting possible. Books also deliver vile and transportation is a convenient tool for terrorists and government crackdowns. Society has found ways to limit these negative effects, and is finding more ways to limit as new mutations crop up. Deleterious attributes are also observable in the new technologies. The mobile communication instrumentation and ubiquitous information availability has changed us. It takes us away from longer range careful thinking and articulating, as well as how we react to any information thanks to the trust that the written word has instilled in us. Financial markets, social interactions, and even our family interactions show consequences of this access ubiquity that has arisen from the nanoscale. Drones as no-blood-on-my-hand machines of death are now ubiquitous in conflicts trivializing death and human destiny exploiting the technology.

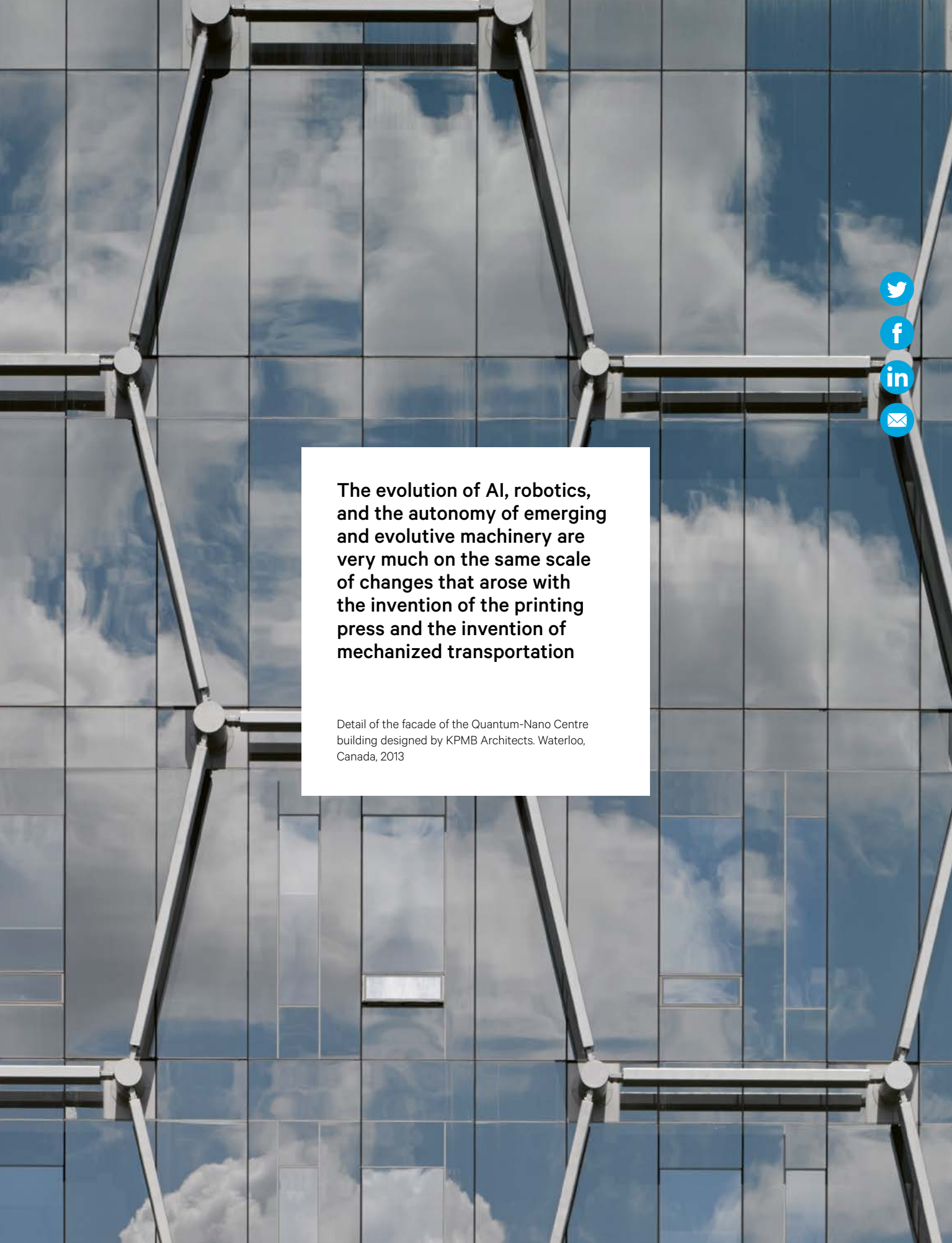
**Quantum computation is an area that has made tremendous progress over the last decade. Numerous difficult problems, cryptography having been the earliest rationale for pursuing this approach, but also many more practical interesting problems—understanding molecules and molecular interactions, working toward further complexity of them including into drug discovery—become potentially solvable**

This evolution in artificial intelligence, robotics, autonomy, the merging of emergent and evolutionary machinery is raising profound questions and society is struggling to grasp even a rudimentary understanding of them so that it can rationally tackle from both a philosophical and a scientific perspective many of the problems that will most certainly arise, just as they did with earlier moments of technology-seeded inflection.

I would like to explore this segue by looking into the coming era's possibilities and integrating thoughts that encompass where this comes from and where it could lead to by looking at some unifying thoughts across the physical and natural world integrated by nanotechnology.

The second law of thermodynamics, first formulated by Sadi Carnot in 1824, says that in an isolated system entropy always increases. Entropy, in a simplistic sense, is a measure of randomness. The law says that if there is absent movement of energy and matter in a system, the system will evolve toward complete randomness. It will exhaust itself of any capability. A thermodynamic equilibrium is this state of maximum randomness. Initially empirical, it now has concrete statistical foundations, and is the reason why natural processes tend to go in one direction, as also for the arrow of time. In contrast, our world—natural and what we have created—has a lot of organization together with a capability to do really interesting things. Atoms assemble themselves into molecules. Molecules become complex. Very complex molecules such as the ribosomes—consisting of RNA and protein parts—become a controller and perform transcription and messenger functions essential to creating proteins. Cells appear—thousands in variety in the human—organs form, and grow to end up in the diversity of nature's creations. A living emergent machine appeared through a hierarchy





**The evolution of AI, robotics,  
and the autonomy of emerging  
and evolutive machinery are  
very much on the same scale  
of changes that arose with  
the invention of the printing  
press and the invention of  
mechanized transportation**

Detail of the facade of the Quantum-Nano Centre  
building designed by KPMB Architects. Waterloo,  
Canada, 2013





**AI coupled to robotics—  
the physical use of this  
information—is also slowly  
moving from factories to  
human-centric applications**

Foreground view of glass reinforced with a nanolayer. The photograph was taken at the P Glass factory in Moscow



of this organized coupling. Our social system, computing or governmental or finance systems, are also such a machinery: parts coupling to other parts under an organized orthodoxy that has numerous capabilities. The second law's vector toward thermodynamic randomness has been overcome due to flow of matter and energy and an interesting diversity has come about.

This is the appearance of complexity taking the form of hierarchy for which Herbert Simon has an interesting parable in the essay "The architecture of complexity."<sup>3</sup> It involves Hora and Tempus, the watchmakers. Both make watches with a thousand little individual parts. Tempus made his watches by putting all the parts together in one go, but if interrupted, for example, by the phone, he had to reassemble it from all these one thousand parts. The more the customers liked the watch, the more they called, the more Tempus fell behind. Hora's watches were also just as good. But he made them using a hierarchy of subassemblies. The first group of subassemblies used ten parts each. Then ten such subassemblies were used to build a bigger subassembly, and so on. Proud competitors in the beginning, Tempus ended up working for Hora. If there is a one-in-a-hundred chance of being interrupted during the assembling process, Tempus, on average, had to spend four thousand times more time than Hora to assemble a watch. He had to start all over again from the beginning. Hora had to do this only part of the way. Hierarchy made Hora's success at this rudimentary complex system possible.

Thermodynamics' lesson in this parable, as also in a generalization to our physical and natural world, is that the complexity arose from the assembling of parts which in general may be random. The likelihood of building up from a small number of elements coming together to obtain a stable assembly, initially simple, but progressively more complex with the building up of hierarchy, is larger than for the coming together of a large number of elements. And new properties emerged in the end watch and possibly in the intermediate subassemblies. Of course, this is a very simplistic description subject to many objections, but an organizational structure appeared due to the flow of energy and parts into the system and the existence of stable intermediate forms that had a lowering of entropy. If there is enough time, nature too will build hierarchies based on stable intermediate forms that it discovers. By doing so, it acquires negentropy (a lowering—negative—of entropy from its state of exhaustion, that is, the maximum). This is the story of the appearance of life.

## **Our social system, computing or governmental or finance systems, are also such a machinery: parts coupling to other parts under an organized orthodoxy that has numerous capabilities**

In his 1967 book *The Ghost in the Machine*,<sup>4</sup>—the third time that this phrase has appeared for us—Arthur Koestler calls this process the "Janus effect." Nodes of this hierarchy are like the Roman god, whose one face is toward the dependent part and the other toward the apex. This link and proliferation of these links with their unusual properties are crucial to the emergent properties of the whole.

Thermodynamics, however, places additional constraints and these arise in the confluence of energy, entropy, and errors. Any complex system consists of a large number of hierarchical subassemblies. A system using nanoscale objects is subject to these constraints of the build-





Porior millenem exceaue corat et rempore officatemqui  
con nonsedit aut que repel et eic to iust, consequid  
quundis doluptur, ullat hicilitio eum recte est ut aut lab  
id ex et dolupta tioria deni re, oditis inverio nsent, susam  
remoles diaestem voloreh endaect inciam, conse pro







ing process. If errors are to be reduced, the reduction of entropy must be large. But, such a reduction process requires the use of more energy. Heat will arise in this process. To be sustainable—avoiding overheating, even as energy flow keeps the system working—requires high-energy efficiency, limits to the amount of energy transformed, and keeping a lid on errors. Having to deploy so much energy at each step makes the system enormously energy hungry. The computation and communication disciplines suffer from this thermodynamic consequence. The giant turbines that convert energies to electric form—the mechanical motion of the blades to the flow of current across voltages—need to be incredibly efficient so that only a few percent—if that—of that energy is lost to the volume. And a turbine is not really that complex a system. Computation and communication have not yet learned this lesson.

Nature has found its own clever way around this problem. Consider the complex biological machine that is the human. The building or replacement processes occur individually in very small volumes—nanoscale—and are happening in parallel at Avogadro number scale. All this transcription, messaging, protein and cell generation, and so on, requires energy and nature works around the error problem by introducing mechanisms for self-repair. Very low energy ( $10\text{--}100 k_B T$ , the  $k_B T$  being a good measure of the order of energy in a thermal motion) breaks and forms these bonds. Errors scale exponentially with these pre-factors. At  $10 k_B T$  energy, one in 100,000 building steps, for example, for each unzipping and copying step, will have an error, so errors must be detected, the building machine must step back to the previous known good state and rebuild that bond, a bit like that phone call to Tempus and Hora, but causing Hora to restart only from an intermediate state. Nature manages to do this pretty well. The human body recycles a body weight of ATP (adenosine triphosphate)—the molecule for energy transformation—every day so that chemical synthesis, nerve impulse propagation, and muscle contraction can happen. Checking and repairing made this complexity under energy constraint possible.

**What would take nature's complex system many many generations, such as through plant breeding, or of living species, can be achieved in one or few generations. An emergent-evolution machine has been born with human intervention. But now it can live on its own. New members of the natural kingdom can be engineered**

How is all this discussion related to the subject of nanotechnology with its masquerading ghost?

I would like to illustrate this by bringing together two technologies of this decade: that of CRISPR and the rise of artificial intelligence.

CRISPR is a gene-editing method that uses a protein (Cas9, a naturally occurring enzyme) and specific guide RNAs to disrupt host genes and to insert sequences of interest. A guide RNA that is complementary to a foreign DNA sequence makes it possible for Cas9 to unwind the sequence's DNA helix, create a double-stranded break, and then the repair enzyme puts it back together by placing a desired experimental DNA. The guide RNA sequence is relatively inexpensive to design, efficiency is high, and the protein injectable. Multiple genes can be mutated in one step. This process can be practiced in a cell's nucleus, in stem cells,

in embryos, and extracellularly. It is a gene-editing tool that allows the engineering of the genome. Thermodynamics teaches us that errors will always occur; if one puts in a lot more energy, the likelihood is less, and this is why low-energy systems need self-correction mechanisms. Even then, rare errors can still pass through. CRISPR helps with that for the genomic errors in the natural world. A defective Janus—an error in the Hora-like way of building the system—can be fixed.

Drug discovery, genetic disease prevention, heart diseases, blood conditions, modifying plants<sup>5</sup> for a change of properties—tomatoes that are jointless; a working example is the joint where the tomato attaches to the plant that gets a signal to die and let go when a tomato is ripe—and others all become possible. What would take nature's complex system many many generations, such as through plant breeding, or of living species, can be achieved in one or few generations. An emergent-evolution machine has been born with human intervention. But now it can live on its own. New members of the natural kingdom can be engineered.

The second technology is that of machine learning now evolving toward an artificial intelligence. As electronic devices have become smaller and smaller, and new architectures—including those with rudimentary self-test and self-repair—evolved, with ever-higher densities and integration levels, we now have supercomputing resources local at our desk as well as in the cloud. This enormous data sorting and manipulating power is now sufficient for the programmed algorithms to discover patterns, find connections between them, and also launch other tasks that ask questions to help test the validity of inferences and probe for more data where enough is not known. The human view of information is of it representing compact associations. While the picture of a dog contains a lot of data in the pixels, and one can even associate a scientific information measure—a negentropy of sorts—to it, the information in a human context is that these associations when viewing the picture will be of it being a dog, more specifically a golden retriever that looks well looked after, but also that golden retrievers are prone to cancer due to selective breeding, and so on. Quite a bit of this information can be associated with the dog as nuggets of wisdom. Even that dogs and cats do not always get along together, or that dogs became domesticated—this too by selective breeding—during the human as hunter-gatherer phase. As we go farther and farther away from the dog as a pattern—dogs evolved from wolves—a physical machine's ability gets weaker and weaker unless the connection of very different domains of information can be put together as knowledge. For a human this is less of a problem. If this factual information of wolf-dog existed as a content in the machine, it is fine, but if one were interested in finding this connection, it will require asking questions, genetic tracing, pattern matching, and taking other steps before the connection will be established. The human learning and knowledge creation happens through hypotheses, looking at results of experiments and debates before ultimately a consensus appears. This capability, because machines can ask and research for information, is now within the reach of artificial intelligence, although connecting two what were previously separate domains—space and time as in relativity—is certainly not, at least not yet. So, when one reads reports of artificial intelligence performing a preliminary assessment of diseases, helping design molecules, and building experiments for the designer molecules, these are fairly complex pattern-matching tasks where data and information is being connected.

With quantum computing also making enormous progress, there is another range of possibilities of powerful import. Because of the exceptional and different computing capability that quantum computation achieves (entanglement is subsumed), making connections that





exist, but are not apparent classically, can be potentially made latent quantum-mechanically. This potentially implies that the nature of associations and connections that the human sees, as Einstein saw in relativity, and classical machine learning only sees if it has already seen it, may be unraveled by the quantum computing machinery since they exist in the entangled data and are not classically manifest.

We can now connect our thermodynamic thoughts to this discussion.

CRISPR as a gene-modification tool came about by replacing nature's randomized trials process with human intervention, because scientists knew how bacteria's immune response fights a foreign viral DNA. CRISPR is this producing of two strands of short RNA (the guide RNA), that then go and form a complex with the Cas9 enzyme that then targets and cuts out and thus disables the viral DNA. Because of the way Cas9 binds to DNAs, it can distinguish between bacterial DNA and viral DNA. So, in this ability lies the memory that can continue to correct future threats. The threats are the errors. From the domain of bacterial viral infection, we have now made connections to the vast domain of natural processes that depend on the genome because a process that nature had shown as being a stable has now been connected to other places, which nature may have discovered, but perhaps not evolved to in the Wallace-Darwin sense. Both the energy constraint and the error correction and stability constraint of thermodynamics were satisfied in the physical practice of CRISPR in the natural world. Human beings can now apply this to multitudes of places where nature does not because it has either not figured it out or because it figured out that it was inappropriate.

**Drug discovery, genetic disease prevention, heart diseases, blood conditions, modifying plants for a change of properties—tomatoes that are jointless; a working example is the joint where the tomato attaches to the plant that gets a signal to die and let go when a tomato is ripe—and others all become possible with CRISPR**

Now look at the evolution of computing to artificial intelligence in our physical technology and its capabilities as well as shortcomings. Physical electronics devices getting to smaller-length scales means that much that happens at the quantum scale becomes important. Tunneling is one such phenomena, and electrons tunneling because voltages exist, not because data is being manipulated, is just a production of heat and not doing anything useful. To get low errors in these physical systems—without yet a general ability to correct errors at the individual step stage—means that one cannot reduce the energy being employed to manipulate the data. So, the amount of energy that is consumed by a biological complex system to make a useful transformation is many orders of magnitude lower than that of the physical complex system. The human brain is like a 20 W bulb in its energy consumption. The physical systems used in the artificial intelligence world are hundreds of times more power consuming if sitting on a desktop and another factor of hundreds more if resident in a data center. So, this artificial intelligence capability has arrived with an enormous increase in power consumption because of the thermodynamics of the way the complex system is implemented.

Human thinking and the human brain is a frontier whose little pieces we may understand, but not the complex assembly that makes it such a marvel. So, the tale of "Blind men and the elephant" should keep serving us as a caution. But there are a number of characteristics that



scientists—psychologists, behaviorists, neuroscientists—do seem to agree with through their multifaceted experiments. The behaviorist Daniel Kahneman<sup>6</sup> classifies human thinking as fast and slow. Fast is a System 1 thinking that is fast, heuristic, instinctive, and emotional. In making quick judgments, one maps other contexts to the problem one is faced with. This dependence on heuristics causes bias, and therefore systematic errors. Slow is a System 2 thinking that is deliberative and analytic, hence slow. The human brain, and those of only a few other mammals, is capable of imagining various pasts and future “worlds” to develop situational inferences and responses. The neuroscientist Robert Sapolsky<sup>7</sup> discusses the origin of aggression and gratification in terms of the events in the brain. I associate these mostly with the fast since they tend to be instinctive. Functionally, the brain can be partitioned into three parts. Layer 1—common to the natural kingdom—is the inner core that is quite autonomic that keeps the body on an even keel by regulating. Layer 2, more recent in evolution, expanded in mammals and is proportionally the most advanced in humans. Layer 3—the neocortex, the upper surface—is the youngest in development (a few hundred million years<sup>8</sup>), relatively very recent, with cognition, abstraction, contemplation, sensory processing as its fortes. Layer 1 also works under orders from Layer 2 through the hypothalamus. Layer 3 can send signals to Layer 2 to then direct Layer 1. The amygdala is a limbic structure—an interlayer—below the cortex. Aggression is mediated by the amygdala, the frontal cortex, and the cortical dopamine system. The dopaminergic system—dopamine generation in various parts of the brain—is activated in anticipation of reward, so this complex system is for this particular aspect a trainable system—a Pavlovian example—where learning can be introduced by reinforcing and inhibitive training experiences. A profound example of the failure of this machinery was Ulrike Meinhof, of the Baader-Meinhof Red Army Faction that was formed in 1968 in Germany. Earlier, as a journalist, she had had a brain tumor removed in 1962. Her autopsy in 1976 showed a tumor and surgical scar tissue impinging on the amygdala, where social contexts, anxieties, ambiguities, and so on are tackled to form that response of aggression. Many of the proclivities of a human depend on this Layer 3 to Layer 2 machinery and the Januses embedded in them. Pernicious errors, if not corrected, the style of thinking and decision-making as in intelligence, matters profoundly. This is true for nature and will be true for artificial intelligence.

## **Nanotechnology has now nearly eliminated the gulf between the physical and the biological domain. The last decades of the twentieth century and those that followed brought about numerous nanotechnology tools that made a more fundamental understanding of complex biological processes possible**

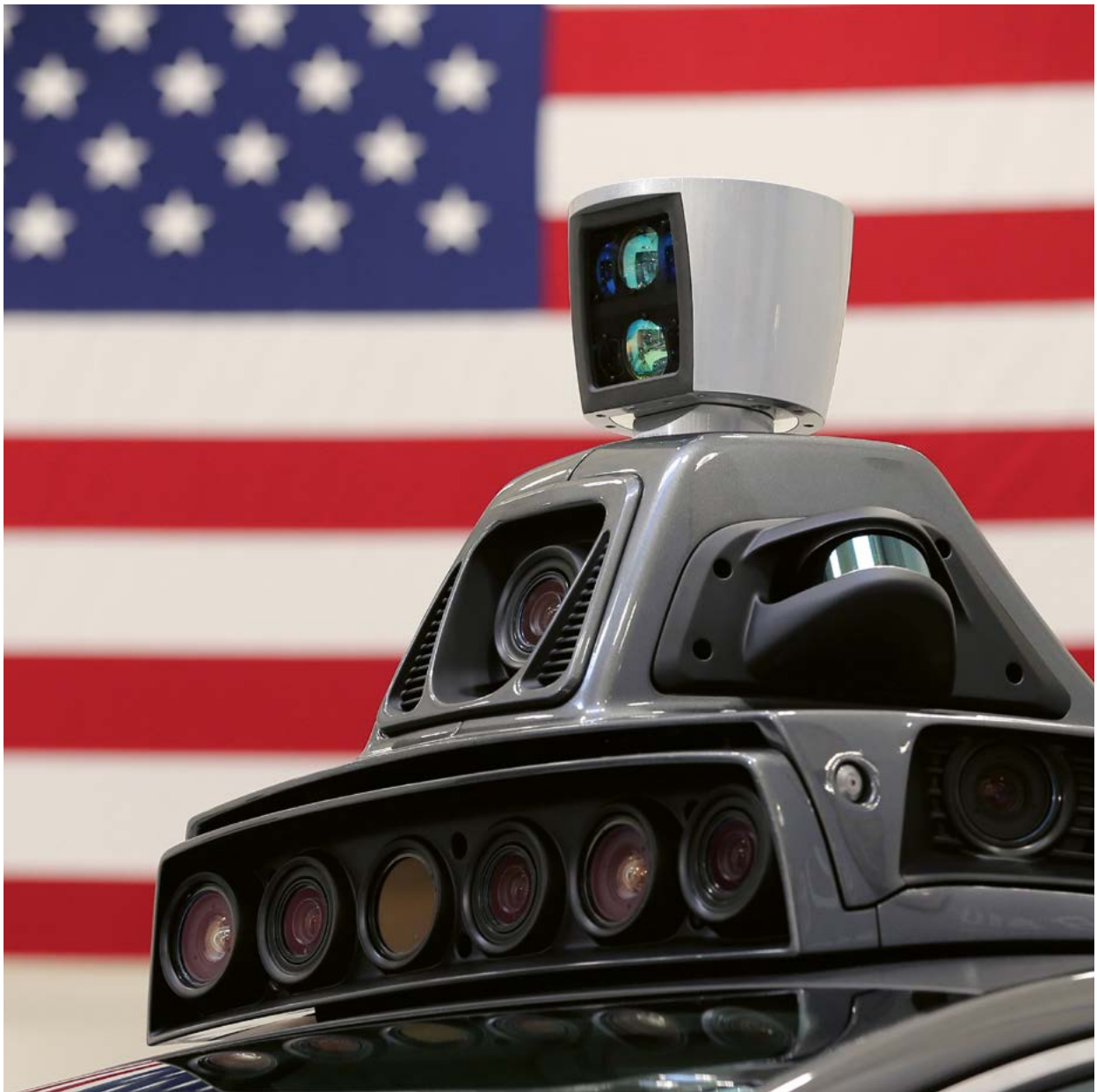
With this background of the current state and related discussion, one should now turn to look at the future. Technology, when used judiciously and for the greater good of the societal collective, has been a unique human contribution to nature. The technological capabilities over the last nearly two hundred years, since the invention of the steam engine, have been compounding at an incredible rate.

In this milieu, in recent decades, nanotechnology has now nearly eliminated the gulf between the physical and the biological domain. The last decades of the twentieth century





Techno-camera and radar system of an UBER Ford Fusion autonomous vehicle during a technology demonstration of driverless cars in Pittsburgh, Pennsylvania, on September 13, 2016





and those that followed brought about numerous nanotechnology tools that made a more fundamental understanding of complex biological processes possible. In turn, this has led to the incredible manipulation capabilities that now exist that sidestep the natural ways of evolution. Artificial intelligence brings in brain-like cognitive capability, except that this capability can access incredibly more data than any human can keep and often it is in the hands of corporations.

The story of technology tells us that the nineteenth-century Industrial Revolution occurred in mechanics, which served to automate physical labor. I am going to call this new nanotechnology-enabled change with the merging of the biological and physical an *existential revolution* since it impinges directly on us as beings.

This existential revolution and the changes and disruption it is likely to cause are of a kind and at such a scale that the traditional long delay of society exercising control—it takes many decades, even societal upheavals such as those arising in Marxism following the Industrial Revolution—will just not work. Much care, foresight, and judgment is needed, otherwise we will witness consequences far worse than the loss of species that we are currently witnessing due to the unbridled expansion of humanity and the need for resources for his expansion.

There are plenty of illustrations pointing toward this, some more deeply connected than others.

**The story of technology tells us that the nineteenth-century Industrial Revolution occurred in mechanics, which served to automate physical labor. I am going to call this new nanotechnology-enabled change with the merging of the biological and physical an *existential revolution* since it impinges directly on us as beings**

Between 1965 and 2015, the primary energy supply of the planet has increased from 40 PWh to 135 PWh (petawatt hour as the unit) over a whole year. This is an increase in excess of a factor of three over this fifty-year period with the predominant source of energy being oil, coal and gas. During this same period, the yearly electrical energy movement in the US alone went from about 8 PWh to nearly 42 PWh, a five-fold increase. The United States alone now consumes in the form of electric energy the total energy that the entire world consumed about fifty years ago. US consumption is a quarter of the energy consumption of the world even though its population is less than one twentieth. Much of this energy-consumption increase has come from the proliferation of computation and communication. Each of that smartphone's functioning requires a semi-refrigerator of network and switching apparatus. Each of the web searches or advertising-streamed courtesy of machine-learning algorithms is another semi-refrigerator of computational resources. Global warming is at its heart a problem arising in this electricity consumption, which in turn comes from this information edifice and its thermodynamic origins.

As the cognitive aspect of the machine ability improves, and robotics—autonomous cars being a primary trend currently—proliferates, will people live farther and farther away from work because cars have been transformed into offices on wheels, or will it lead to an efficient usage of cars where cars pick up people and drop off people in set places and so fewer cars



are needed because they are being used efficiently and perhaps only exist as a service that is owned by the community or a business? If it is the latter, this is a technological solution around a major problem that has appeared in our society a hundred plus years after the invention of the combustion engine. If it is the former, there is now another increase of energy consumption brought about by technology. Existentially insidious is the continuing expansion of the “medium is the message” because technology companies with the computational platforms and the reach are really media companies pushing advertising, thoughts, and lifestyles leveraging their access to the individual’s societal interaction stream.

In the biological disruption, humans have not quite objected to genetically modified agriculture, which is a rudimentary genome modification. Improved nutritional value, pest and stress resistance, and other properties have been welcomed, particularly in countries such as the US and in the Third World. Not so well understood is the crossbreeding and introduction of foreign transgenes into nature, nor is the effect of such artificial constructs that did not arise through the evolutionary process of nature on other natural species. Increased resistance also goes together with invasiveness, leading to reduced diversity. Bt-corn fertilizes other crops and becomes a vector for cross-contamination of genes in the Mandeleevian style. Bt-corn also affects other insects, a prominent example being of Monarch butterflies. So, just like the antibiotics-resistant bacteria that we now face, especially in the hospitals of the Third World with tuberculosis being the most insidious, and the major hospitals of the US, we will have deleterious consequences from the new engineering of biology. We have few capabilities for visualizing these possible threats, so this subject is worth dwelling on.

## **As the cognitive aspect of the machine ability improves, and robotics—autonomous cars being a primary trend currently—proliferates, will people live farther and farther away from work because cars have been transformed into offices on wheels, or will it lead to an efficient usage of cars**

What does the introduction of a precisely programmed artificial change into the natural world do? That is, a world in which evolution takes place via a selection based on small and inherited variations that increase the ability to compete, survive, and reproduce by natural random interactions of which only specific ones will continue. It will lead to further continuation of the natural evolution process, but in which a nonnatural entity has now been introduced. Nature evolves slowly on a long-term time scale. This evolving minimizes harmful characteristics all the way from the basic molecular level to the complex system level at each of the steps of the assembling. While the artificial change has a specific characteristic desired by the human, it may not and will not for many instances have the characteristics that nature would have chosen because of the optimization inherent in the natural evolution process. The compression of the time scale and the introduction of specific nonnatural programming means that the chances of errors will be large. We see the deleterious consequences of such errors in systems all the time in our world’s complex systems. An error of placing the wrong person in the hierarchy of an organization leads to the failures of such organizations. An error of placing the wrong part in the complex system, such as in Hora’s watch, will stop its working. A specific function may be achieved through the deliberate making of a choice, but a global function may be lost.



CRISPR makes the consequences of this genetic modification of the natural world much more profound, because it can make programmed change over multiple genes possible, not just in crops but in all the different inhabitants—living and nonliving. Multiple changes affected simultaneously compounds the risks. A hamburger-sized tomato that has flavor, takes long to rot, and grows on a jointless plant may be very desirable to McDonalds, the farmer, and the consumer, but it is extremely risky with the risks not really even ascertainable about what the jumping of genes across species could possibly lead to. This is precisely what the thermodynamic constraints on nature's processes—energies, error rates, error-correcting, generational mutations, and self-selection—has mitigated for billions of years.

A short step from this is the potential for programming traits of our own offspring. Who does not want brighter and more beautiful children? A hundred years ago, eugenics was enormously popular in the countries of the West—John Maynard Keynes, Teddy Roosevelt, Bertrand Russell, Bernard Shaw, Winston Churchill were all adherents—and it did not take long for this path's culmination in Adolf Hitler. Among the pockets of high incidence of Asperger's syndrome is the Silicon Valley, where many of the high-technology practitioners—a group with specific and similar traits and thus reduced diversity—reside. That autism exists as a problem here should not be a cause for surprise. Genes serve multitudes of purposes, and various traits arise in their coexistence. This requires diversity from which the multitudes of traits—of disease, disposition, and others—that humans acquire. Program this, and we really cannot tell what deleterious outcomes will arise and one will not see these for generations.

Another scaling of this approach is the putting together of CRISPR with artificial intelligence as an emergent-evolution fusion. With enough data, one could ask for a certain set of characteristics in this group, the algorithms design CRISPR experiments to achieve it, and behold, we have engineered a new human, or even a new species.

**It is the challenge to society to bring about a civilized order to all this. Society needs to find ways to function so that self-repair becomes possible at each stage of the complex system building. Neither top-down nor bottom-up suffice, the repairing must exist at each stage in how society functions and also in how the technology works**

As Leo Rosten once said: "When you don't know where a road leads, it sure as hell will take you there." This is the existential paradox.

So, I return to the dilemma outlined at beginning of this paper. In the first half of the twentieth century, when we learned to create chemicals for specific purposes with similar motives, for example, pesticides, even as agriculture output increased—a desirable outcome—we also faced Rachel Carson's *Silent Spring*.<sup>9</sup> With time, we found ways to mitigate it, but this problem still festers at low intensity.

It is the challenge to society to bring about a civilized order to all this. Society needs to find ways to function so that self-repair becomes possible at each stage of the complex system building. Neither top-down nor bottom-up suffice, the repairing must exist at each stage in how society functions and also in how the technology works.



This complex world that we now have has this existential dilemma in the emergent-evolutionary fusion.

We have moved beyond Descartes and Ryle, and the great phenomenologists.

The ghost exists in this machinery. Ghosts can be angelic and ghosts can be demonic. Angels will benefit us tremendously, just as transportation has for mobility, drugs have for health, communication has for our friendships and family, and computing has in helping build the marvels and technologies that we employ every day. Humans must understand and shape the technology. If we can build such systems with clear and well-thought out understanding of what is not acceptable as a human effort, keep notions of relevance and provenance, and make the systems reliable, then the world's education, health, agriculture, finance, and all the other domains that are essential to being civilized humans as citizens of this planet can all be integrated together for the greater good of all.

But, if not dealt with judiciously and at the planet scale with nature's kingdom at its heart, the human story could very well follow the memorable line from Arthur Clarke's *2010: Odyssey Two* book: "The phantom was gone; only the motes of dancing dust were left, resuming their random patterns in the air."



## Notes

1. Sandip Tiwari, "Paradise lost? Paradise regained? Nanotechnology, man and machine," in *There's a Future: Visions for a Better World*, OpenMind/BBVA, 2012).
2. Artificial intelligence was a word coined by John McCarthy in the late 1950s as the coming together of software and hardware—logic and logical processing—to a human intelligence-like capability. Norbert Weiner's cybernetics at that time referred to the coming together of techniques—control to information to patterns to statistics—in building intelligent systems. Machine learning is certainly closer to cybernetics at the moment.
3. Herbert A. Simon, "The architecture of complexity," *Proceedings of the American Philosophical Society*, 106(6): 467–482 (1962).
4. Arthur Koestler, *The Ghost in the Machine*, Hutchinson & Co., London, 1967. This book was a follow on to *The Act of Creation*, where he explored creativity. This follow on explored the human destructive behavior, and posited these actions on the rapid growth of human brain with destructive impulses subduing logical construction. Modern work in neurosciences posits considerable human behavior on the flow between the top layer—the frontal cortex—and lower layers which took a much much longer time to develop.
5. "Crispr can speed up nature—and change how we grow food" <https://www.wired.com/story/crispr-tomato-mutant-future-of-food/> (accessed on September 19, 2018).
6. Daniel Kahneman, *Thinking Fast and Slow*, Farrar, Straus and Giroux, New York, 2011.
7. Robert M. Sapolsky, *Behave. The Biology of Humans at Our Best and Worst*, Bodley Head, London, 2017.
8. Arthur Koestler blames this short span of evolution of the cortex for the error that the aggression trait represents. It is interesting that humans dislike the "wrong kind" of aggression, but in the "right context" we admire it. This aggression takes many forms and it can be tied to thought, emotion, and action; offensive and defensive; impulsive and premeditated; emotional, cold-blooded, and hot-blooded; for pleasure; and as a displacement, when being an affected party at the hands of somebody else, we take it out on another weaker person. Humans are masters of aggression, well beyond the rest of the mammal kingdom, because of this developed brain. This short span must also have some bearing on Nietzsche's two moralities: a "master morality" that values strength, beauty, courage, and success, and a "slave" morality that values kindness, empathy and sympathy.
9. Rachel Carson, *Silent Spring*, Houghton Mifflin, Boston, 1962.



**Joanna J. Bryson**  
University of Bath

Joanna J. Bryson is a transdisciplinary researcher on the structure and dynamics of human- and animal-like intelligence. Her research, covering topics from artificial intelligence, through autonomy and robot ethics, and on to human cooperation, has appeared in venues ranging from a reddit to *Science*. She holds degrees in Psychology from Chicago and Edinburgh, and in Artificial Intelligence from Edinburgh and MIT. She has additional professional research experience from Princeton, Oxford, Harvard, and LEGO, and technical experience in Chicago's financial industry, and international management consultancy. Bryson is presently a Reader (associate professor) at the University of Bath.

Recommended book: *WTF? What's the Future and Why It's Up to Us*, Tim O'Reilly, Harper Business, 2017.

**Artificial intelligence (AI) is a technical term referring to artifacts used to detect contexts or to effect actions in response to detected contexts. Our capacity to build such artifacts has been increasing, and with it the impact they have on our society. This article first documents the social and economic changes brought about by our use of AI, particularly but not exclusively focusing on the decade since the 2007 advent of smartphones, which contribute substantially to “big data” and therefore the efficacy of machine learning. It then projects from this political, economic, and personal challenges confronting humanity in the near future, including policy recommendations. Overall, AI is not as unusual a technology as expected, but this very lack of expected form may have exposed us to a significantly increased urgency concerning familiar challenges. In particular, the identity and autonomy of both individuals and nations is challenged by the increased accessibility of knowledge.**

## 1. Introduction



The past decade, and particularly the past few years, have been transformative for artificial intelligence (AI) not so much in terms of what we can do with this technology as what we *are* doing with it. Some place the advent of this era to 2007, with the introduction of smartphones. As I detail below, at its most essential, intelligence is just intelligence, whether artifact or animal. It is a form of computation, and as such a transformation of information. The cornucopia of deeply personal information that resulted from the willful tethering of a huge portion of society to the Internet has allowed us to pass immense explicit and implicit knowledge from human culture via human brains into digital form. Here we can not only use it to operate with human-like competence, but also produce further knowledge and behavior by means of machine-based computation.

For decades—even prior to the inception of the term—AI has aroused both fear and excitement as humanity contemplates creating machines in our image. This expectation that *intelligent* artifacts should by necessity be *human-like* artifacts blinded most of us to the important fact that we have been achieving AI for some time. While the breakthroughs in surpassing human ability at human pursuits, such as chess (Hsu, 2002), Go (Silver et al., 2016), and translation (Wu et al., 2016), make headlines, AI has been a standard part of the industrial repertoire since at least the 1980s. Then production-rule or “expert” systems became a standard technology for checking circuit boards and detecting credit card fraud (Liao, 2005). Similarly, machine-learning (ML) strategies like genetic algorithms have long been used for intractable computational problems, such as scheduling, and neural networks not only to model and understand human learning, but also for basic industrial control and monitoring (Widrow et al., 1994). In the 1990s, probabilistic and Bayesian methods revolutionized ML and opened the door to some of the most pervasive AI technologies now available: search through massive troves of data (Bishop, 1995). This search capacity included the ability to do semantic analysis of raw text, astonishingly enabling Web users to find the documents they seek out of trillions of Web pages just by typing only a few words (Lowe, 2001; Bullinaria and Levy, 2007).

## **Today, AI breakthroughs in surpassing human ability in certain activities make headlines, but AI has been a standard part of the industrial repertoire since at least the 1980s**

This capacity to use AI for discovery has been extended not only by the massive increase of digital data and available computing power, but also by innovations in AI and ML algorithms. We are now searching photographs, videos, and audio (Barrett et al., 2016; Wu et al., 2016). We can translate, transcribe, read lips, read emotions (including lying), forge signatures and other handwriting, and forge video (Hancock et al., 2007; Eyben et al., 2013; Assael et al., 2016; Haines et al., 2016; Reed et al., 2016; Vincent, 2016; Schuller et al., 2016; Sartori et al., 2016; Thies et al., 2016; Deng et al., 2017; Chung and Zisserman, 2017). Critically, we can forge real-time audio/video during live transmissions, allowing us to choose the words millions “witness,” particularly for celebrities such as politicians for whom there is already a great deal of data for composing accurate models (Thies et al., 2016; Suwajanakorn et al., 2017). At the time of this writing, there is increasing evidence that the outcomes of the 2016 US presidential election and UK referendum on EU membership were both altered by the





AI detection and targeting of “swing voters” via their public social-media use (Cadwalladr, 2017a,b; ICO, 2018), not to mention the AI-augmented tools used in cyberhacking (Brundage et al., 2018). AI is here now, available to and benefiting us all. But its consequences for our social order are not only not understood, but have until recently barely even yet been the subject of study (Bryson, 2015). Yet now too, with advances in robotics, AI is entering our physical spaces in the form of autonomous vehicles, weapons, drones, and domestic devices, including “smart speakers” (really, microphones) and even games consoles (Jia et al., 2016). We are becoming surrounded by—even embedded in—pervasive automated perception, analysis, and increasingly action.

What have been and will be the impacts of pervasive synthetic intelligence? How can society regulate the way technology alters our lives? In this article, I begin by presenting a clean, clear set of definitions for the relevant terminology. I then review concerns and suggested remediations with respect to technology. Finally, I make expert though unproven recommendations concerning the value of individual human lives, as individual human capacities come increasingly under threat of redundancy to automation.

## 2. Definitions

The following definitions are not universally used, but derive from a well-established AI text Winston (1984), as well as from the study of biological intelligence (Barrows, 2000, attributed to Romanes, 1883). They are selected for clarity of communication at least local to this chapter, about the existing and potential impacts of intelligence, particularly in machines. *Intelligence* is the capacity to do the right thing at the right time, in a context where doing nothing (making no change in behavior) would be worse. Intelligence then requires:

- the capacity to perceive *contexts* for action;
- the capacity to *act*;
- the capacity to *associate* contexts to actions.

By this definition, plants are intelligent (Trewavas, 2005). So is a thermostat (McCarthy, 1983; Touretzky, 1988). They can perceive and respond to context: for example, plants to the direction of light, thermostats to temperature. We further discriminate a system as being *cognitive* if it is able to modify its intelligence, something plants and at least mechanical thermostats cannot do. Cognitive systems are able to learn new contexts, actions, and/or associations between these. This comes closer to the conventional definition of “intelligent.”

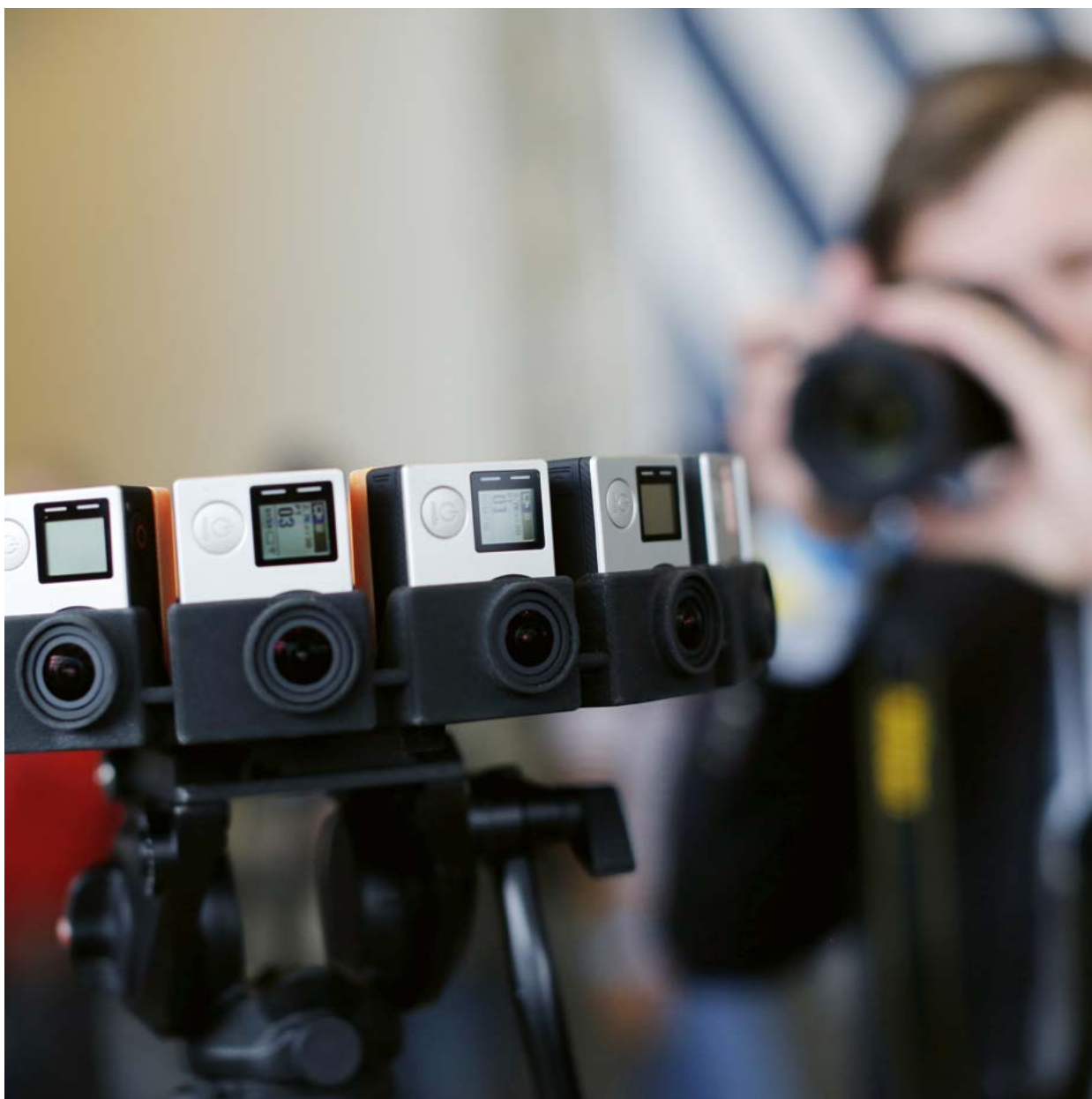
***Intelligence* is the capacity to do the right thing at the right time, in a context where doing nothing (making no change in behavior) would be worse**

Intelligence as I defined it here is a strict subset of *computation*, the transformation of information. Note that computation is a physical process, it is not maths. It takes time, space, and energy. Intelligence is the subset of computation that transforms a context into action.

*Artificial* intelligence (AI), by convention, is a term used to describe (typically digital) artifacts that extend any of the capacities related to natural intelligence. So, for example,



Presentation of a GoPro device at the I/O developers congress in San Francisco in May 2015. This device includes sixteen cameras that can be used with Google's *Jump* software to provide 360° vision





machine vision, speech recognition, pattern recognition, and fixed (unlearning) production systems are all considered examples of AI, with algorithms that can be found in standard AI textbooks (Russell and Norvig, 2009). These can also all be seen as forms of computation, even if their outputs are not conventionally seen as action. If we embrace, though, the lessons of embodied robotics (see below), then we might extend this definition to include as AI *any* artifact that extends our own capacities to perceive and act. Although this would be an unusual definition, it might also give us a firmer grip on the sorts of changes AI brings to our society, by allowing us to examine a longer history of technological interventions.

*Machine learning* (ML) is any means of programming AI that requires not only conventional hand coding, but also a component of automated generalization over presented data by means of accumulating statistics on that data (Murphy, 2012; Erickson et al., 2017). Often, but not necessarily, ML comes down to seeking regularities in data that are associated with categories of interest, including appropriate opportunities for particular actions. ML is also often used to capture associations, and can be used to acquire new action skills, for example from demonstration (Huang et al., 2016).

**Intelligence is a strict subset of *computation*, the transformation of information. Note that computation is a physical process, it is not maths. It takes time, space, and energy. Intelligence is the subset of computation that transforms a context into action**

Note that all ML still involves a hand-programmed component. The mere conceptualization or discovery of an algorithm never leads to a machine capable of sensing or acting springing spontaneously into existence. All AI is by definition an *artifact*, brought into being by deliberate human acts. Something must be built and designed to connect some data source to some representation before any learning can occur. All intelligent systems have an *architecture*, a layout through which energy and information flows, and nearly always including locations where some information is retained, termed *memory*. The design of this architecture is called *systems engineering*; it is at this point that a system's safety and validity should be established. Contrary to some outrageous but distressingly frequent claims, AI safety is not a new field. Systems engineering in fact predates computers (Schlager, 1956), and has always been a principal component of computer-science education. AI has long been integrated into software, as documented in the introduction, so there is a long history of it being engineered in safe ways (for example, Bryson, 2003; Chessell and Smith, 2013).

*Robots* are artifacts that sense and act in the physical world, and in real time. By this definition a smartphone is a (domestic) robot. It has not only microphones but also a variety of proprioceptive sensors that allow it to know when its orientation is changing or it is falling. Its range of actions includes intervening with its user and transmitting information including instructions to other devices. The same is true of many game consoles and digital home assistants—"smart speakers"/microphones like Google Home, Amazon's Echo (Alexa), or Microsoft's Cortana.

*Autonomy* is technically the capacity to act as an individual (Armstrong and Read, 1995; Cooke, 1999). So, for example, a country loses its autonomy either if its institutions collapse so that only its citizens' individual actions have efficacy, or if its institutions come under the



influence of other agencies or governments to such an extent that again its own government has no impact on its course of actions. Of course, either extreme is very unusual. In fact, for social animals like humans autonomy is never absolute (Gilbert et al., 2012). Our individual intelligence determines many of our actions, but some cells may become cancerous in pursuit of their own goals counter to our overall well-being (Hanahan and Weinberg, 2011). Similarly, we fully expect a family, place of work, or government, to have impact on our actions. We also experience far more social influence implicitly than we are ever aware of (Devos and Banaji, 2003). Nevertheless, we are viewed as autonomous because there is an extent to which our own individual intelligence also influences our behavior. A technical system able to sense the world and select an action specific to its present context is therefore called “autonomous” even though its actions will ultimately be determined by some combination of the designers that constructed its intelligence and its operators. Operators may influence AI in real time, and will necessarily influence it in advance by setting parameters of its operation, including when and where it operates, if at all. As discussed earlier, designers call the system into existence and determine its capacities, particularly what information it has access to and what actions it can take. Even if a designer chooses to introduce an element of chance, such as dependence on the present environment or a random-number generator into the control of an AI system, that inclusion is still the deliberate choice of the designer.

## **AI safety is not a new field. Systems engineering in fact predates computers (Schlager, 1956), and has always been a principal component of computer-science education**

### **3. Concerns about AI and Society**

AI is core to some of the most successful companies in history in terms of market capitalization—Apple, Alphabet, Microsoft, and Amazon. Along with Information and Communication Technology (ICT) more generally, AI has revolutionized the ease with which people from all over the world can access knowledge, credit, and other benefits of contemporary global society. Such access has helped lead to massive reduction of global inequality and extreme poverty, for example by allowing farmers to know fair prices, best crops, and giving them access to accurate weather predictions (Aker and Mbiti, 2010).

AI is the beneficiary of decades of regulatory policy: research and deployment has so far been largely up-regulated with massive government and other capital investment (Miguel and Casado, 2016; Technology Council Committee on Technology, 2016; Brundage and Bryson, 2017). Although much of the emphasis of later parts of this paper focuses on possible motivations for, or mechanisms of, regulatory restriction on AI, it should be recognized that:

1. any such AI policies should and basically always will be developed and implemented in the light of the importance of respecting the positive impacts of technology as well;<sup>2</sup>
2. no one is talking about introducing regulation to AI. AI already exists in a regulatory framework (Brundage and Bryson, 2017; O'Reilly, 2017); what we are discussing is whether that framework needs optimizing;



3. regulation has so far mostly been entirely constructive, with governments providing vast resources to companies and universities developing AI. Even where regulation constrains, informed and well-designed constraint can lead to more sustainable and even faster growth.

Having said this, academics, technologists, and the general public have raised a number of concerns that may indicate a need for down-regulation or constraint. Smith (2018), president of Microsoft, recently asserted:

[Intelligent<sup>3</sup>] technology raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression. These issues heighten responsibility for tech companies that create these products. In our view, they also call for thoughtful government regulation and for the development of norms around acceptable uses. In a democratic republic, there is no substitute for decision-making by our elected representatives regarding the issues that require the balancing of public safety with the essence of our democratic freedoms.

In this section I categorize perceived risks by the sort of policy requirements they are likely to generate. I also make recommendations about whether these are nonproblems, problems of ICT or technology more generally, or problems special to AI, and in each case what the remedy may be.

**3.1. Artificial General Intelligence (AGI) and Superintelligence** I start with some of the most sensational claims—that as artificial intelligence increases to the point that it surpasses human abilities, it may come to take control over our resources and outcompete our species, leading to human extinction. As mentioned in Section 1, AI is already superhuman in many domains. We can already do arithmetic better, play chess and Go better, transcribe speech better, read lips better, remember more things for longer, and indeed be faster and stronger with machines than unaided. While these capacities have disrupted human lives including employment (see below), they have in no way led to machine ambition.

Some claim that the lack of machine ambition, or indeed domination, is because the forms of AI generated so far are not sufficiently general. The term *artificial general intelligence* (AGI) is used to describe two things: AI capable of learning anything without limits, and human-like AI. These two meanings of AGI are generally conflated, but such conflation is incoherent, since in fact human intelligence has significant limitations. Understanding the limitations of human intelligence is informative because they relate also to the limits of AI.

**We can already do arithmetic better, play chess and Go better, transcribe speech better, read lips better, remember more things for longer, and indeed be faster and stronger with machines than unaided**

Limitations on human intelligence derive from two causes: combinatorics and bias. The first, combinatorics, is a universal problem affecting all computation and therefore all natural and artificial intelligence: *combinatorics* (Sipser, 2005). If an agent is capable of one hundred actions, then it is capable of 10,000 two-step plans. Since humans are capable of far more than





one hundred different actions and perform far more than two actions even in a day, we can see that the space of possible strategies is inconceivably vast, and cannot be easily conquered by any scale of intelligence (Wolpert, 1996b).

However, computer science has demonstrated that some ways of exploring such vast spaces are more effective than others, at least for specific purposes (Wolpert, 1996a). Most relevantly to intelligence, concurrent search by many processors simultaneously can be effective provided that the problem space can be split between them, and that a solution once found can be both recognized and communicated (Grama, 2003). The reason human technology is so much more advanced than other species' is because we are far more effective at this strategy of concurrent search, due to our unique capacity to share advances or "good tricks" via language (Dennett, 2013; Bryson, 2008, 2015; van Schaik et al., 2017). Our culture's increasing pace of change is in part due to the unprecedented number of individuals with good health and education connected together by ICT, but also to our augmentation of our search via machine computing. Our increasing capacities for AI and artifactual computation more generally increase further our potential rate of exploration; quantum computation could potentially accelerate these far further (Williams, 2010). However, note that these advantages do not come for free. Doing two computations at once may double the speed of the computation if the task was perfectly divisible, but it certainly doubles the amount of space and energy needed to do the computation. Quantum computing is concurrent in space as well as time, but its energy costs are so far unknown, and very likely to be exorbitant.

## **Our culture's increasing pace of change is in part due to the unprecedented number of individuals with good health and education connected together by ICT, but also to our augmentation of our search via machine computing**

Much of the recent immense growth of AI has been largely due to improved capacities to "mine" using ML the existing discoveries of humanity and nature more generally (Moeslund and Granum, 2001; Calinon et al., 2010; Caliskan et al., 2017). The outcomes of some of our previous computation are stored in our culture, and biological evolution can also be thought of as a massive parallel search, where the outcomes are collated very inefficiently, only as fast as the best genes manage to reproduce themselves. We can expect this strategy of mining past solutions to soon plateau, when artificial and human intelligence come to be sharing the same, though still-expanding, boundary of extant knowledge.

The second source of limitations on human intelligence, which I called "bias" above, are those special to our species. Given the problems of combinatorics, all species only explore a tiny subset of possible solutions, and in ML such focus is called *bias*. The exact nature of any biological intelligence is part of its evolutionary niche, and is unlikely to be shared even by other biological species except to the extent that they have similar survival requirements and strategies (Laland et al., 2000). Thus, we share many of our cognitive attributes—including perception and action capacities, and, importantly, motivations—with other apes. Yet we also have specialist motivations and capacities reflecting our highly social nature (Stoddart, 1990). No amount of intelligence in itself necessitates social competitiveness, neither does it demand the desire to be accepted by an ingroup, to dominate an outgroup, nor to achieve recognition within an ingroup. These are motivations that underlie human cooperation and



competition that result from our evolutionary history (Mace, 1998; Lamba and Mace, 2012; Jordan et al., 2016; Bryson et al., 2017); further, they vary even among humans (Van Lange et al., 1997; Herrmann et al., 2008; Sylwester et al., 2017). For humans, social organizations easily varied to suit a politico-economic context are a significant survival mechanism (Stewart et al., 2018).

None of this is necessary—and much of it is even incoherent—from the perspective of an artifact. Artifacts are definitionally designed by human intent, not directly by evolution. With these intentional acts of authored human creation<sup>4</sup> comes not only human responsibility, but an entirely different landscape of potential rewards and design constraints (Bryson et al., 2017; Bryson, 2018).

Given all of the above, AGI is obviously a myth—in fact, two orthogonal myths:

1. no amount of natural or artificial intelligence will be able to solve all problems;
2. even extremely powerful AI is exceedingly unlikely to be very human-like, because it will embody an entirely different set of motivations and reward functions.

These assertions, however, do not protect us from another, related concern. *Superintelligence* is a term used to describe the situation when a cognitive system not only learns, but learns how to learn. Here again there are two component issues. First, at this point an intelligence should be able to rapidly snowball to such an extent that it would be incomprehensible to ordinary human examination. Second, even if the intelligence was carefully designed to have goals aligned with human needs, it might develop for itself unanticipated subgoals that are not. For example, a chess-playing robot might learn to shoot the people that deprive it of sufficient resources to improve its game play by switching it off at night, or a filing robot might turn the planet into paperclips in order to ensure all potential papers can be adequately ordered (Bostrom, 2012).

These two examples are ludicrous if we remember that all AI systems are designed and a matter of human responsibility. No one has ever made a chess program that represents information concerning any resources not on the chessboard (with the possible exception of time), nor with the capacity to fire a gun. The choice of capacities and components of a computer system is again part of its architecture. As I mentioned earlier, the systems engineering of architecture is an important component to extant AI safety, and as I will say below (Section 4.3), it can also be an important means for regulating AI.

However, the concept of superintelligence itself is not ludicrous; it is clear that systems that learn to learn can and do experience exponential growth. The mistake made by futurists concerned with superintelligence is to think that this situation is only a possible future. In fact, it is an excellent description of human culture over the last 10,000 years, since the innovation of writing (Haberl et al., 2007; Barnosky, 2008). The augmentation of human intelligence with technology has indeed resulted in a system that has not been carefully designed and results in unintended consequences. Some of these consequences are very hazardous, such as global warming and the reduction of species diversity. List and Pettit (2011) make a similar point when they call human organizations such as corporations or governments “AI.”

As I mentioned, I will return to the importance of architecture and design again, but it is worth emphasizing once more here the necessity of such biases and limits. Robots make it particularly apparent that behavior depends not only on computational capacities but also on other system attributes, such as physical capacities. Digital manipulation, such as typing or playing the flute, is just not an option for either a smartphone or a snake, however intelligent. Motivations are similar. Unless we design a system to have anthropomorphic goals, social



perception, and social behavior capacities, we are not going to see it learning to produce anthropomorphic social behavior, such as seeking to dominate conversations, corporations, or countries. If corporations do show these characteristics, it is because of the expression of the human components of their organization, and also because of the undisciplined, evolutionary means by which they accrete size and power. From this example we can see that it is possible for an AI system—at the very least by the List and Pettit (2011) argument—to express superintelligence, which implies that such intelligent systems should be regulated to avoid this.

## **The concept of superintelligence itself is not ludicrous; it is clear that systems that learn to learn can and do experience exponential growth. The mistake made by futurists concerned with superintelligence is to think that this situation is only a possible future**

From the above I conclude that the problem of superintelligence is real but not special to AI; it is, rather, one our cultures already face. AI is, however, now a contributing factor to our capacity to excel, but this may also lead us to learn to better self-regulate—that is, govern—as it has several times in the past (Milanovic, 2016; Scheidel, 2017). Even were AGI to be true and the biological metaphor of AI competing by natural selection to be sound, there is no real reason to believe that we would be extinguished by AI. We have not extinguished the many species (particularly microbial) on which we ourselves directly depend. Considering unintended consequences of the exponentially increasing intelligence of our entire socio-technical system (rather than AI on its own) does, however, lead us to more substantial concerns.

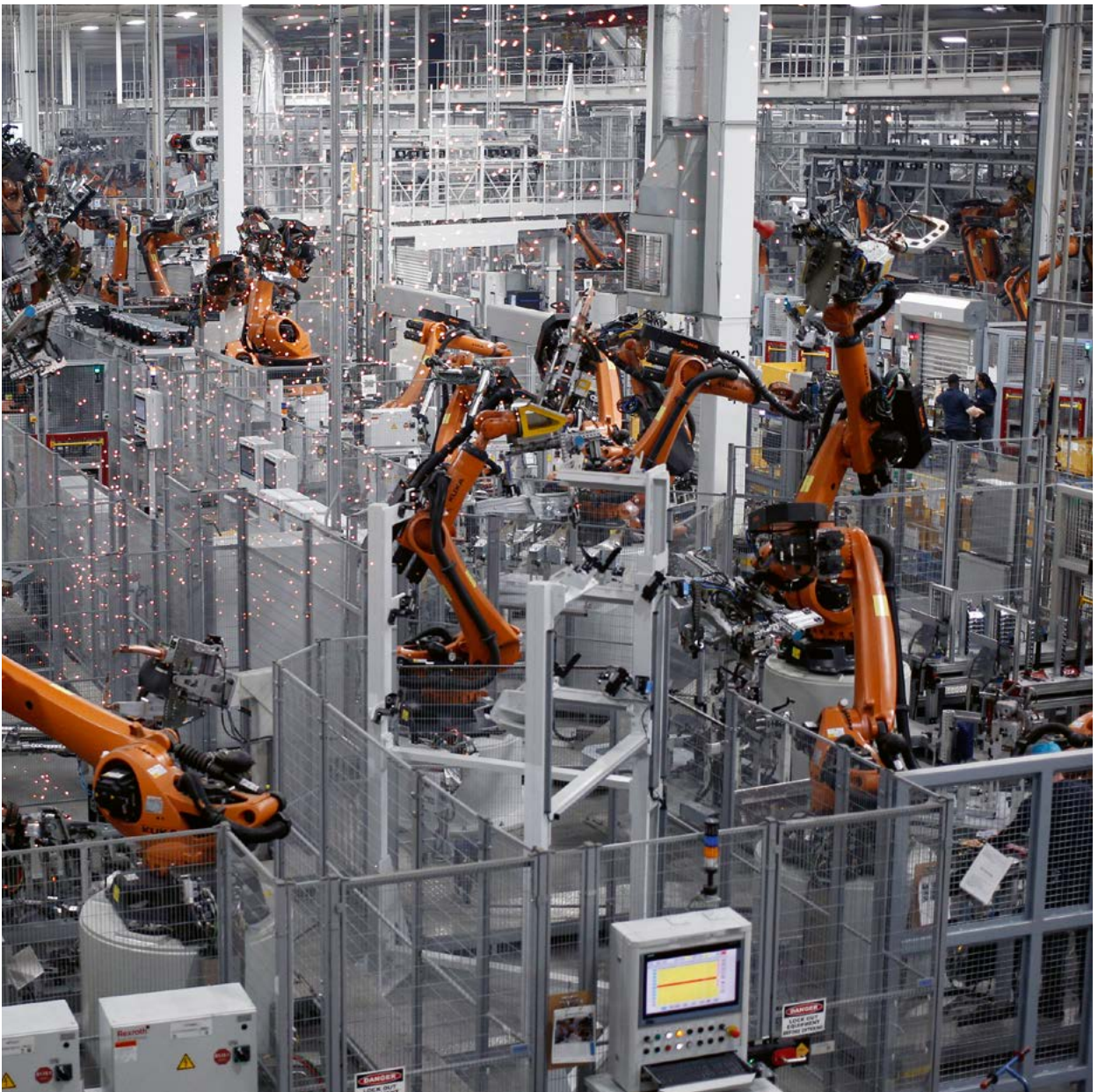
**3.2. Inequality and Employment** For centuries there have been significant concerns about the displacement of workers by technology (Autor, 2015). There is no question that new technologies do disrupt communities, families, and lives, but also that historically the majority of this disruption has been for the better (Pinker, 2012). In general, lifespans are longer and infant mortality lower than ever before, and these indicators are good measures of contentedness in humans, as low infant mortality in particular is well associated with political stability (King and Zeng, 2001).

However, some disruption does lead to political upheaval, and has been recently hypothesized to associate with the rise of AI. Income (and presumably wealth) inequality is highly correlated with political polarization (McCarty et al., 2016). Political polarization is defined by the inability of political parties to cooperate in democratic governance, but periods of polarization are also characterized by increases in identity politics and political extremism. Political polarization and income inequality covary but either can lead the other; the causal factors underlying the relationship are not well understood (Stewart et al., 2018). What is known is that the last time these measures were as high as they are now (at least in the OECD) was immediately before and after World War I. Unfortunately, it took decades of policy innovation, a global financial crisis, and a second world war before inequality and polarization were radically reduced and stabilized in the period 1945–78 (Scheidel, 2017), though note that in some countries such as the USA and UK the second shock of the financial crisis was enough.





Robots welding components at the Bayerische Motoren Werke, A.G. (BMW) assembly plant in Greer, South Carolina, May 2018





Fortunately, we now know how to redress this situation—redistribution lowers inequality. After World War II, when tax rates were around fifty percent, modern welfare states were built or finalized, transnational wealth extraction was blocked (Bullough, 2018), and both income inequality and political polarization were kept low for over twenty years. During this time, wages also kept pace with productivity (Mishel, 2012). However, some time around 1978 wages plateaued, and both inequality and political polarization began rising, again in the OECD.<sup>5</sup> The question is what caused this to happen. There are many theories, but given the starkness of the shift on many metrics it looks more like a change in policy than of technology. This could reflect geopolitical changes of the time—it could signal, for example, the point at which economically influential members of the OECD detected the coming end of the Cold War, and shifted away from policies designed to combat the threat of Communist uprisings.

## **Income inequality and political polarization might be the result of the rise of AI, but the fact that similar political and economic trends occurred in the late 1800s indicates that this is not a special concern of any one technology**

Regardless of the causes, with respect to AI, the fact that similar political and economic trends occurred in the late 1800s again means that this is not a special concern of any one technology. While, as mentioned, there is so far no consensus on causes, in ongoing research, I with other authors<sup>6</sup> are exploring the idea that some technologies reduce costs that had traditionally maintained diversity in the economic system. For example, when transport costs are high, one may choose to use a nearby provider rather than finding the global best provider for a particular good. Similarly, lack of information transparency or scaling capacity may result in a more diverse use of providers. Technical innovations (including in business processes) may overcome these costs and allow relatively few companies to dominate. Examples from the late nineteenth century might include the use of oil, rail, and telegraph, and the improvement of shipping and newspaper delivery.

Where a few providers receive all the business, they will also receive all of the wealth. Governance is a primary mechanism of redistribution (Landau, 2016), thus revolutions in technology may require subsequent revolutions in governance in order to reclaim equilibrium (Stewart et al., 2018). The welfare state could be one such example (Scheidel, 2017). We will return to discussing the possible need for innovations in governance below (Section 4).

To return to AI or more likely ICT, even if these technologies are not unique in contributing to inequality and political polarization, they may well be the principal component technologies presently doing so. Further, the public and policy attention currently directed toward AI may afford opportunities to both study and address the core causes of inequality and polarization, particularly if AI is seen as a crisis (Tepperman, 2016). Nevertheless, it is worth visiting one hypothesized consequence of polarization in particular. An increase in identity politics may lead to the increased use of beliefs to signal ingroup status or affiliation (Iyengar et al., 2012; Newman et al., 2014), which would unfortunately decrease their proportional use to predict or describe the world—that is, to reflect facts. Thus, ironically, the age of information may not universally be the age of knowledge, but rather also an age of disinformation.<sup>7</sup>

This reliance on beliefs as ingroup indicators may influence another worrying trait about contemporary politics: loss of faith in experts. While occasionally motivated by the irre-



sponsible use or even abuse of position by some experts, in general losing access to experts' views is a disaster. The combinatorial explosion of knowledge mentioned in Section 3.1 also means that no one, however intelligent, can master in their lifetime all human knowledge. If society ignores the stores of expertise it has built up—often through taxpayer-funding of higher education—it sets itself at a considerable disadvantage.

These concerns about the nature and causes of “truthiness” in what should be the information age lead also to our next set of concerns, about the use of personal information.

**3.3. Privacy, Personal Liberty, and Autonomy** When we consider the impact of AI on individual behavior, we now come to a place where ICT more clearly has a unique impact. There have long been periods of domestic spying which have been associated with everything from prejudiced skew in opportunities to pogroms. However, ICT is now allowing us to keep long-term records on anyone who produces storable data—for example, anyone with bills, contracts, digital devices, or a credit history, not to mention any public writing and social-media use. That is, essentially, everyone.

It is not only the storage and accessibility of digital records that changes our society; it is the fact that these can be searched using algorithms for pattern recognition. We have lost the default assumption of anonymity by obscurity (Selinger and Hartzog, 2017). We are to some extent all celebrities now: any one of us can be identified by strangers, whether by facial-recognition software or data mining of shopping or social-media habits (Pasquale, 2015). These may indicate not just our identity but our political or economic predispositions, and what strategies might be effective for changing these (Cadwalladr, 2017a,b). ML allows us to discover new patterns and regularities of which we may have had no previous conception. For example, that word choice or even handwriting pressure on a digital stylus can indicate emotional state, including whether someone is lying (Hancock et al., 2007; Bandyopadhyay and Hazra, 2017), or a pattern of social-media use can predict personality categories, political preferences, and even life outcomes (Youyou et al., 2015).

**It is not only the storage and accessibility of digital records that changes our society; it is the fact that these can be searched using algorithms for pattern recognition. We have lost the default assumption of anonymity by obscurity**

Machine learning has enabled near-human and even superhuman abilities in transcribing speech from voice, recognizing emotions from audio or video recordings, as well as forging handwriting or video (Valstar and Pantic, 2012; Griffin et al., 2013; Eyben et al., 2013; Klein-smith and Bianchi-Berthouze, 2013; Hofmann et al., 2014; Haines et al., 2016; Reed et al., 2016; Vincent, 2016; Thies et al., 2016; Deng et al., 2017). The better a model we have of what people are likely to do, the less information we need to predict what an individual will do next (Bishop, 2006; Youyou et al., 2015). This principle allows forgery by taking a model of a person's writing or voice, combining it with a stream of text, and producing a “prediction” or transcript of how that person would likely write or say that text (Haines et al., 2016; Reed et al., 2016). The same principle might allow political strategists to identify which voters are likely to be persuaded if not to change party affiliation, at least to increase or decrease their probability of turning out to vote, and then to apply resources to persuade them to do so. Such





a strategy has been alleged to have impacted significantly on recent elections in the UK and USA (Cadwalladr, 2017a,b; ICO, 2018); if so, they were almost certainly tested and deployed earlier in other elections less carefully watched.

Individuals in our society might then reasonably fear the dissemination of their actions or beliefs for two reasons: first because it makes them easier to predict and therefore manipulate; and second because it exposes them to persecution by those who do not approve of their beliefs. Such persecution could range from bullying by individuals, through to missed career or other organizational opportunities, and on to in some unstable (or at least unethical) societies, imprisonment or even death at the hands of the state. The problem with such fears is not only that the stress of bearing them is itself noxious, but also that in inhibiting personal liberty and free expression we reduce the number of ideas disseminated to society as a whole, and therefore limit our ability to innovate (Mill, 1859; Price, 1972). Responding to both opportunities and challenges requires creativity and free thinking at every level of society.

**3.4. Corporate Autonomy, Revenue, and Liability** These considerations of personal autonomy lead directly to the final set of concerns I describe here, which is not one frequently mentioned. Theoretical biology tells us that where there is greater communication, there is a higher probability of cooperation (Roughgarden et al., 2006). While cooperation is often wonderful, it can also be thought of as essentially moving some portion of autonomy from the individual to a group (Bryson, 2015). Let us recall from the Section 2 definitions that the extent of autonomy an entity has is the extent to which it determines its own actions. Individual and group autonomy must to some extent trade off, though there are means of organizing groups that offer more or less liberty for their constituent parts. Thus, the limits on personal liberty just described may be a very natural outcome of introducing greater capacity for communication. Here once more, I again refer to all of ICT, but AI and ML with their capacity to accelerate the search for both solutions and collaborators are surely a significant component, and possibly game-changing.

One irony here is that many people think that bigger data is necessarily better, but better for what? Basic statistics teaches us that the number of data points we need to make a prediction is limited by the amount of variation in that data, providing only that the data is a true random sample of its population.<sup>8</sup> The extent of data we need for science or medicine may require only a minuscule fraction of a population. However, if we want to spot specific individuals to be controlled, dissuaded, or even promoted, then of course we want to “know all the things.”

But changing the costs and benefits of investment at the group level have more consequences than only privacy and liberty. ICT facilitates blurring the distinction between customer and corporation, or even the definition of an economic transaction. This has so far gone largely unrecognized, though see Perzanowski and Schultz (2016); Frischmann and Selinger (2016). Customers now do real labor for the corporations to whom they give their custom: pricing and bagging groceries, punching data at ATMs for banks, filling in forms for airlines, and so forth (Bryson, 2015). The value of this labor is not directly remunerated—we assume that we receive cheaper products in return, and as such our loss of agency to these corporations might be seen as a form of bartering. They are also not denominated, obscuring the value of this economy. Thus, ICT facilitates a black or at least opaque market that reduces measured income and therefore tax revenue where taxation is based on denominated turnover or income. This problem holds for everyone using Internet services and interfaces, even ignoring the problematic definitions of employment raised by platforms (though see O’Reilly, 2017). Our



improving capacity to derive value and power while avoiding revenue may also help explain the mystery of our supposed static productivity (Brynjolfsson et al., 2017).

This dynamic is most stark in the case of “free” Web services. Clearly, we are receiving information and/or entertainment in exchange for data and/or attention. If we attend to content co-presented with advertisements, we afford the presenters an opportunity to influence our behavior. The same is true for less conventional forms of nudging, for example the postulated political interventions mentioned in Section 3.3. However, these exchanges are only denominated (if at all) in aggregate, and only when the corporation providing such service is valued. Much of the data is even collected on speculation; it may be of no or little value until an innovative use is conceived years later.



## **ICT facilitates blurring the distinction between customer and corporation, or even the definition of an economic transaction. This has so far gone largely unrecognized**

Our increasing failure to be able to denominate revenue at the traditional point—income, or exchange—may be another cause for increasing wealth inequality, as less of the economy is recognized, taxed, and redistributed. An obvious solution would be to tax wealth directly—for example, the market value of a corporation—rather than income. The information age may make it easier to track the distribution of wealth, making this strategy more viable than it has been in the past, particularly relative to the challenges of tracking income, if the latter challenges are indeed increasing as I described. However, it is inadequate if that wealth is then taxed only in the country (often a tax haven) in which the corporation is formally incorporated. Given that we can see the transnational transfer of data and engagement with services, we should in theory be able to disseminate redistribution in proportion to the extent and value of data derived. Enforcing such a system transnationally would require substantial innovations, since ordinarily taxation is run by government, and there is almost definitionally no transnational government. There are, however, international treaties and organized economic areas. Large countries or coordinated economies, such as the European Economic Area, may be able to demand equitable redistribution for their citizens in exchange for the privilege of access to those citizens. China has successfully demonstrated that such access is not necessarily a given, and indeed blocking access can facilitate the development of local competition. Similar strategies are being used by American cities and states against platforms such as Uber and Airbnb.

Taxation of ICT wealth leads me to a proposed distortion of law that is particularly dangerous. In 2016 the European Parliament proposed that AI or robotics might be reasonably taxed as “e-persons.” This is a terrible idea (Bryson et al., 2017). It would allow corporations to automate part of their business process, then break off that piece in such a way as to limit their liabilities for both taxes and legal damages.

The idea of taxing robots has populist appeal for two reasons. First, it seems basic common sense that if robots are “taking our jobs” they should also “pay taxes” and thus support “us” via the welfare state. Second, many people find appealing the idea that we might extend human life—or something more essential about humanity than life—synthetically via AI and/or robotics. Unfortunately, both of these ideas are deeply incoherent, resting on ignorance about the nature of intelligence.



Any AI policy should and basically always will be developed and implemented in the light of the importance of respecting the positive impacts of technology

European Parliament members voting in the Strasbourg, France, chamber in March 2018









As described in Section 3.1 earlier, both of these appeals assume that *intelligent* means in part *human-like*. While there is no question that the word has been used that way culturally, by the definitions presented in Section 2 it is clearly completely false. To address the second concern first, the values, motivations, even the aesthetics of an enculturated ape cannot be meaningfully shared with a device that shares nothing of our embodied physical (“phenomenological”) experience (Bryson, 2008; Claxton, 2015; Dennett, 2017). Nothing we build from metal and silicon will ever share our phenomenology as much as a rat or cow, and few see cows or rats as viable vessels of our posterity.

## **Taxing robots and extending human life via AI are ideas with populist appeal. Unfortunately, both are based on ignorance about the nature of intelligence**

Further, the idea that substantiating a human mind in digital technology—even were that possible—would make it immortal or even increase its lifespan is ludicrous. Digital formats have a mean lifetime of no more than five years (Lawrence et al., 2000; Marshall et al., 2006). The fallacy here is again to mistake computation for a form of mathematics. While mathematics really is pure, eternal, and certain, that is because it is also not real—it is not manifest in the physical world and cannot take actions. Computation in contrast is real. As described earlier, computation takes time, space, and energy (Sipser, 2005). Space is needed for storing state (memory), and there is no permanent way to achieve such storage (Krauss and Starkman, 2000).

To return to the seemingly more practical idea of taxing AI entities, this again overlooks their lack of humanity. In particular, AI is not countable as humans are countable. This criticism holds also for Bill Gates’s support of taxing robots, even though he did not support legal personality (author pool, 2017). There is no equivalent of “horsepower” to measure the number of humans replaced by an algorithm. As just mentioned, in the face of accelerating innovation we can no longer keep track of the value even of transactions including human participants. When a new technology is brought in, we might briefly see how many humans are made redundant, but even this seems to reflect more the current economy than the actual value of labor replaced (Autor, 2015; Ford, 2015). When times are good, a company will retain and retrain experienced employees; when times are bad corporations will take the excuse to reduce headcount. Even if the initial shift in employment were indicative of initial “person-power” replaced, technologies quickly change the economies into which they are inserted, and the values of human labor rapidly change too.

It is essential to remember that artifacts are by definition designed. Within the limits of the laws of physics and computation, we have complete authorship over AI and robotics. This means that developers will be able to evade tax codes in ways inconceivable to legislators used to value based on human labor. The process of decomposing a corporation into automated “e-persons” would enormously magnify the present problems of the over-extension of legal personhood such as the shell corporations used for money laundering. The already restricted sense in which it is sensible to consider corporations to be legal persons would be fully dissolved if there are no humans employed by the synthetic entity (Solaiman, 2017; Bryson et al., 2017).



#### 4. The Next Ten Years: Remediations and Futures



To stress again as at the beginning of Section 3, AI has been and is an incredible force of both economic growth and individual empowerment. We are with it able to know, learn, discover, and do things that would have been inconceivable even fifty years ago. We can walk into a strange city not knowing the language yet find our way and communicate. We can take advantage of education provided by the world's best universities in our own homes, even if we are leading a low-wage existence in a developing economy (Breslow et al., 2013). Even in the developing world, we can use the village smartphone to check the fair prices of various crops, and other useful information like weather predictions, so even subsistence farmers are being drawn out of extreme poverty by ICT. The incredible pace of completion of the Human Genome Project is just one example of how humanity as a whole can benefit from this technology (Adams et al., 1991; Schena et al., 1996).

Nevertheless, the concerns highlighted above need to be addressed. I will here make suggestions about each, beginning with the most recently presented. I will be brief here, since, as usual, knowledge of solutions only follows from identification of problems, and the identifications above are not yet agreed but only proposed. In addition, some means for redressing these issues have already been suggested, but I go into further and different detail here.

**4.1. Employment and Social Stability** I have already in Section 3.4 dismissed the idea that making AI legal persons would address the problems of employment disruption or wealth inequality we are currently experiencing. In fact, e-personhood would almost certainly *increase* inequality by shielding companies and wealthy individuals from liability, at the cost of the ordinary person. We have good evidence now that wealthy individual donors can lead politicians to eccentric, extremist position-taking (Barber, 2016) which can lead to disastrous results when coupled with increasing pressure for political polarization and identity politics. It is also important to realize that not every extremely wealthy individual necessarily reveals the level of their wealth publicly.

### **E-personhood would almost certainly *increase* inequality by shielding companies and wealthy individuals from liability, at the cost of the ordinary person**

In democracies, another correlate of periods of high inequality and high polarization is very close elections, even where candidates might otherwise not seem evenly matched. This, of course, opens the door to (or at least reduces the cost of) manipulation of elections, including by external powers. Person (2018) suggests weak countries may be practicing “subtractive balancing” against stronger ones, by disrupting elections and through them governance abilities and therefore autonomy, in an effort to reduce power differentials in favor of the weaker nation. If individuals or coalitions of individuals are sufficiently wealthy to reduce the efficacy of governments, then states also lose their autonomy, including the stability of their borders.

War, anarchy, and their associated instability is not a situation anyone should really want to be in, though those who presently profit from illegal activity might think otherwise. Everyone benefits from sufficient stability to plan businesses and families. The advent of transnational



A teleprompter shows a “virtual student” at an MIT class being recorded for online courses in April 2013, Cambridge, Massachusetts





corporations has been accompanied by a substantial increase in the number and power of other transnational organizations. These may be welcome if they help coordinate cooperation on transnational interests, but it is important to realize that geography will always be a substantial determiner of many matters of government. How well your neighbor's house is protected from fire, whether their children are vaccinated or well educated, will always affect your quality of life. Fresh water, sewage, clean air, protection from natural disasters, protection from invasion, individual security, access to transport options—local and national governments will continue to play an extremely important role in the indefinite future, even if some functions are offloaded to corporations or transnational governments. As such, they need to be adequately resourced.

I recommended in Section 3.4 that one possible solution to the impact on ICT on inequality is to shift priority from documenting and taxing income to documenting and taxing wealth. The biggest problem with this suggestion may be that it requires redistribution to occur internationally, not just nationally, because the richest corporations per Internet<sup>9</sup> are in only one country, though certainly for those outside China—and increasingly for those inside—their wealth derives from global activity. Handling this situation will require significant policy innovations. Fortunately, it is in the interest of nearly all stakeholders, including leading corporations, to avoid war and other destructive social and economic instability. The World Wars and financial crises of the twentieth century showed that this was especially true for the extremely affluent, who at least economically have the most to lose (Milanovic, 2016; Scheidel, 2017), though of course do not often lose their lives.

I particularly admire the flexible solutions to economic hardship that Germany displayed during the recent recession, where it was possible for corporations to *partially* lay off employees, who then received *partial* welfare and free time (Eichhorst and Marx, 2011, p. 80). This allowed individuals to retrain while maintaining for a prolonged period a standard of living close to their individual norms; it also allowed companies to retain valued employees while they pivoted business direction or just searched for liquidity. This kind of flexibility should be encouraged, with both governments and individuals retaining economic capacity to support themselves through periods of crisis. In fact, sufficient flexibility may prevent periods of high change from being periods of crisis.

## **Weak countries may be practicing “subtractive balancing” against stronger ones, by disrupting elections and through them governance abilities and therefore autonomy, in an effort to reduce power differentials in favor of the weaker nation**

If we can reduce inequality, I believe the problems of employment will also reduce, despite any increase in the pace of change. We are a fabulously wealthy society, and can afford to support individuals at least partially as they retrain. We are also fantastically innovative. If money is circulating in communities, then individuals will find ways to employ each other, and to perform services for each other (Hunter et al., 2001; Autor, 2015). Again, this may already be happening, and could account for the decreased rate of change some authors claim to detect in society (for example, Cowen, 2011). A great number of individuals may continue finding avenues of self- and (where successful) other employment in producing services within their own communities, from the social, such as teaching, policing, journalism, and family services,

to the aesthetic, such as personal, home, and garden decoration, and the provision of food, music, sports, and other communal acts.

The decision about whether such people are able to live good enough lives that they can benefit from the advantages of their society is a matter of economic policy. We would want any family to be able, for example, to afford a week's holiday in the nearest large city, or to have their children experience social mobility, for example getting into the top universities in their chosen area purely based on merit. Of course, we expect in this century universal and free access to health care, and primary and secondary education. People should be able to live with their families but also not need to commute for enormous portions of their day; this requires both distributed employment opportunities and excellent, scalable (and therefore probably public) transportation infrastructure.



**If we can reduce inequality, I believe the problems of employment will also reduce, despite any increase in the pace of change. We are a fabulously wealthy society, and can afford to support individuals at least partially as they retrain**

The level of investment in such infrastructure depends in part on the investment both public and private in taxation, and also on how such wealth is spent. Historically we have in some periods spent a great deal on the destruction of others' infrastructure and repair of one's own due to warfare. Now, even if we avoid open ballistic warfare, we must face the necessity of abandoning old infrastructure that is no longer viable due to climate change, and investing in other locations. Of course, this offers a substantial opportunity for redistribution, particularly into some currently economically depressed cities, as was shown by Roosevelt's New Deal, which substantially reduced inequality in the USA well before World War II (Wright, 1974; McCarty et al., 2016).

I am with those who do not believe the universal basic income is a great mechanism of redistribution, for several reasons. First, many hope to fund it by cutting public services, but these may well be increasingly needed as increasing numbers of people cannot deal with the technical and economic complexities of a world of accelerating change. Second, I have seen far too many standing safely in the middle of the road telling television cameras that "the government has never done anything for me," ignorant of massive investment in their education, security, and infrastructure. I think a basic income would easily become as invisible and taken for granted as trash collection and emergency services apparently are.

But, most importantly, I would prefer redistribution to reinforce the importance of local civic communities, that is, to circulate through employment, whether direct or as freelancers and customers. AI and ICT make it easy to bond with people from all over the world, or indeed with entertaining fantasies employing AI technology that are not actually human. But our neighbors' well-being has enormous impacts on our own and are in many senses shared, through the quality of water, air, education, fire and other emergency services, and of course personal security. The best neighborhoods are connected through knowledge and personal concern, that is, localized friendships.

One effective mechanism of increasing redistribution is just the increase of minimum wages (Lee, 1999; Schmitt, 2013). Even if this is only done for government employees, it has knock-on effects for the rest of employers as they compete for the best people, and, of course,



also gives the advantage of having better motivation for good workers to contribute to society through civil service. Although this mechanism has been attacked for a wide variety of reasons (for example, Meyer, 2016), the evidence seems fairly good for positive impacts overall.



**4.2. Privacy, Liberty, and Innovation** Stepping back to the coupled problems of privacy and individual autonomy, we hit an area for which predictions are more difficult or at least more diverse. It is clear that the era of privacy through obscurity is over, as we now have more information and more means to filter and understand information than ever before, and this is unlikely to be changed by anything short of a global disaster eliminating our digital capacity. Nevertheless, we have long been in the situation of inhabiting spaces where our governments and neighbors could in theory take our private property from us, but seldom do except by contracted agreement such as taxation (Christians, 2009). Can we arrive at a similar level of control over our personal data? Can we have effective privacy and autonomy in the information era? If not, what would be the consequences?

First, it should be said that any approach to defending personal data and protecting citizens from being predicted, manipulated, or outright controlled via their personal data requires strong encryption and cybersecurity—*without* back doors. Every back door in cybersecurity has been exploited by bad actors (Abelson et al., 2015). Weak cybersecurity should be viewed as a significant risk to the AI and digital economy, particularly the Internet of Things (IoT). If intelligent or even just connected devices cannot be trusted, they will and should not be welcome in homes or workplaces (Weber, 2010; Singh et al., 2016).

Many thinkers on the topic of technologically mediated privacy have suggested that data about a person should be seen not as an asset of the person but as *part* of that person—an extension of an individual's identity. As such, personal data cannot be owned by anyone but the person to whom it refers; any other use is by lease or contract which cannot be extended or sold onward without consent (Gates and Matthews, 2014; Crabtree and Mortier, 2015). This would make personal data more like your person; if it and therefore you are violated, you should have recourse to the law. There are a variety of legal and technological innovations being developed in order to pursue this model; however, given both the ease of access to data and the difficulty of proving such access, data may be far more difficult to defend than physical personal property (Rosner, 2014; Jentzsch, 2014). Fortunately, at least some governments have made it part of their job to defend the data interests of their citizens (for example, the GDPR, Albrecht, 2016; Danezis et al., 2014). This is for excellent reasons, since, as described above, there are both political and economic consequences of foreign extraction of data wealth and manipulation of individual political preferences and other behavior based on those individual's social-media profiles.

**Many thinkers on the topic of technologically mediated privacy have suggested that data about a person should be seen not as an asset of the person but as *part* of that person—an extension of an individual's identity**

The best situated entities to defend our privacy are governments, presumably through class-action lawsuits of at least the most egregious examples of violation of personal data. Note that such courses of action may require major innovations of international law or treaties, since



some of the most prominent allegations of manipulation involve electoral outcomes for entire countries. For example, the UK's Brexit vote has in the first two years since the referendum (and before any actual exit of the EU) cost the country £23 billion in lost tax revenue, or £44 million a week (Morales, 2018). As mentioned earlier, the Brexit vote is alleged to have been influenced by known AI algorithms, which have been shown to have been funded through foreign investment (ICO, 2018). Ironically, achieving compensation for such damage would almost certainly require international collaboration.

Unfortunately, governments do not always have their citizens' interests at heart, or, at least, not always all of their citizens' interests. Indeed, globally in the twentieth century, one was far more likely to be killed by one's own government than by any foreign actor (Valentino, 2004). More recently, China has been using the surveillance system that was supposed to keep its citizens safe to destroy the lives and families of over a million of its citizens by placing them in reeducation camps for the crime of even casually expressing their Muslim identity (Human Rights Watch, 2018; Editorial Board, 2018). More generally, if governments fear whistle blowing, dissent, or even just shirk the responsibility for guaranteeing dignity and flourishing for all those in their territory, then they can and often will suppress and even kill those individuals. It is extremely dangerous when a government views governing a group of people within its borders as more cost or trouble than their collective potential value for labor, security, and innovation is worth. Aggravating this grave threat, we have both the promise and the hype of intelligent automation as a new, fully ownable and controllable source of both labor and innovation. The inflated discourse around AI increases the risk that a government will (mis) assess the value of human lives to be lower than the perceived costs of maintaining those lives.

We cannot know the sure outcome of the current trajectories, but where any sort of suppression is to be exercised, we can readily expect that AI and ICT will be the means for monitoring and predicting potential troublemakers. China is alleged to be using face-recognition capacities not only to identify individuals, but to identify their moods and attentional states both in reeducation camps and in ordinary schools. Students and perhaps teachers can be penalized if students do not pay attention, and prisoners can be punished if they do not appear happy to comply with their (re)education. ICT systems able to detect and inform teachers to adjust lectures and material toward students' attention and comprehension are also being pitched for classrooms in the West, and are core to personalized AI instruction outside of conventional classrooms. Presumably similar systems are also being developed and probably applied for other sorts of work (for example, Levy, 2015).

## **The inflated discourse around AI increases the risk that a government will (mis)assess the value of human lives to be lower than the perceived costs of maintaining those lives**

If we allow such trends to carry on, we can expect societies that are safer—or at least more peaceful on the streets—more homogenous, less innovative, and less diverse. More people have the means and economic wherewithal to move between countries now than ever before, so we might hope that countries that truly produce the best quality of life including governance and individual protection will be attractors to those who care about personal liberty. We may also hope that with the combined power of those immigrants and their extant citizens' labor and innovation, these countries may come to be able to protect not only themselves but others. We have already seen the EU do such protection by setting standards



of AI ethics such as the GDPR, and of course the United Nations is working with instruments such as the Paris Agreement to protect us all from climate change. In such well-governed and flourishing societies, we would expect to see perhaps an increase rather than a decrease in present levels of liberty, as we come to recognize the problems arising from the surveillance we already practice, for example in micromanaging our children's personal time (Lee et al., 2010; Bryson, 2015).

Unfortunately for this optimistic vision of pools of well-being spreading from well-governed countries, in practice technology is increasingly being used or being threatened to be used for blocking any cross-border migration except by the most elite (Miller, 2017). Besides genocide and mass killings, another historic trend often observed in wars and political revolutions (for example, Nazi-occupied Poland, cold-war Czechoslovakia, the Iranian Revolution, Stalin's USSR, Cambodia under the Khmer Rouge, China's Cultural Revolution, present-day Saudi Arabia) is the displacement or even execution of not only dissidents but all and any intelligentsia. The effort to maintain control is often seen as requiring the elimination of any potential innovative leadership, even though precisely such leadership may be necessary to keep people healthy and therefore civil society stable (King and Zeng, 2001), not to mention maintaining the technological progress necessary to stay abreast in any arms race (Yan, 2006). Such movements tend to fail only after protracted suffering, and often only after having persisted long enough to make clear the damage of their policies to the country's international competitiveness. AI makes the identification and isolation of any such targeted group—or even individuals with target attitudes—spectacularly easy. Only if we are able to also innovate protections against corrupt, selfish, or otherwise dangerous governance can we protect ourselves from losing the diversity and liberty of our societies, and therefore the security of us all.

Again, the capacity for AI to be used for good governance leading to fairer and stronger societies is very real and widely being developed. For example, AI is used to reduce financial crime, fraud, and money laundering, protecting individuals, societies, and governments from undue and illegal influence (Ngai et al., 2011). This is sensible and a part of ordinary contractual understandings of the duties of financial service providers and, indeed, governments. It may also be ethical for citizens to be “nudged” by their devices or other agencies into behaviors they have consciously chosen and explicitly expressed a desire for, such as exercise or sleep regimes. But it is important to recognize the massively increased threats of both explicit duress and implicit misleading that accompanies the massive increase of knowledge and therefore power that derive from AI. AI therefore increases the urgency of investment in research and development in the humanities and social sciences, particularly the political sciences and sociology. I therefore now turn to the problem of regulating AI.

**4.3. Individual, Corporate, and Regulatory Responsibility for AI** To begin the discussion of responsibility in an age of AI, I want to return briefly to emphasize again the role of design and architectures on AI. Again, perhaps because of the misidentification of *intelligent* with *human*, I have sometimes heard even domain experts from influential organizations claim that one or another trait of AI is inevitable. There is no aspect of AI more inevitable than slavery or the hereditary rights of monarchy. Of course, both of these still persist in some places, but despite their economic benefits to those formerly in power, they have been largely eradicated. Similarly, we can regulate at least legal commercial products to mandate safe or at least transparent architectures (Boden et al., 2011). We can require—as again the European Commission recently has—that decisions taken by machines be traceable and explainable (Goodman and Flaxman, 2016).



Maintaining human accountability for AI systems does not have to mean that we must (or can) account for the value of every weight in a machine-learned neural network, or the impact of every individual instance of data used in training. Not only is this impractical, it is not the standard or means by which we currently hold organizations accountable. A company is not responsible for the synaptic organization of its accounts' brains; it is responsible for the state of its accounts. Introducing AI into a corporate or governance process actually changes little with respect to responsibility. We still need to be able to characterize our systems well enough to recognize whether they are behaving as intended (Liu et al., 2017). This is doable, and it should be encouraged (Bryson and Theodorou, 2019).

Encouraging responsibility entails ensuring we continue maintaining accountability (Santoni de Sio and van den Hoven, 2018). One simple way to do this is to educate governments and prosecutors that software systems have much the same liability issues as any other manufactured artifact—if they are misused, it is the fault of the owner; if they cause harm when being appropriately used, they are at fault and the manufacturer is likely liable unless they can prove due diligence and exceptional circumstance. The mere fact that part of the system is autonomous does not alter this fact, just as a bank can be held accountable for errors generated by its accountants or customers where their bank's systems should have caught or constrained such errors. There are certainly challenges here, particularly because so many applications of AI technology are in transnational contexts, but organizations such as the EU, UN, and OECD are looking to be able to coordinate the efforts of nations to protect their citizens.

Of course, AI systems are not exactly like more deterministic systems, but exaggerating the consequences of those differences creates problems. Bad ideas can be hidden by the “smoke and mirrors” of the confusion generated by AI around identity and moral agency (Bryson, 2018). One concerning trend in AI governance is the trend for *value alignment* as a solution to difficult questions of science generally, and AI ethics in particular. The idea is that we should ensure that society leads and approves of where science or technology go (Soares and Fallenstein, 2014). This may sound very safe and democratic, but it is perhaps better seen as populist. Speaking first to science: science is a principal mechanism enabling society to accurately *perceive* its context. In contrast, *governance* is how a society chooses between potential actions. Popular sentiment cannot determine what is true about nature; it can only determine what policies are easiest to deploy. To limit a society's capacity to perceive to only the things it wants to know would be to blind that society (Caliskan et al., 2017, see the final discussion). Similarly, the outcomes of policy are highly influenced by public sentiment, but certainly not determined by it. Asking the public what it wants from AI is like asking them which science-fiction film they would most like to see realized—there is no guarantee they will choose one that is feasible, let alone truly desirable in protracted detail. While the public must through government determine its economic and political priorities, actual progress is almost never achieved by referendum. Rather, governance almost always comes down to informed negotiations between a limited number of expert negotiators, supported by a larger but still limited number of domain experts.

Even given the vast resources available through exploiting computation and AI, it is likely that human negotiators will always be the best determiners of policy. This is partly because we as citizens can identify with human representatives, thus establishing trust and investment in the negotiated outcomes. But more importantly, human representatives can be held to account and persuaded in ways that AI never can be. We cannot intentionally design systems to be as centered on social outcomes as human or indeed any social animal's intelligence has evolved to be. We cannot do this by design, because design by its nature is a decomposable



process, whereas evolution has repeatedly discovered that concern for social standing must be an inextricable part of an individual's intelligence for a species reliant on social strategies for its survival. Thus our entire system of justice relies on dissuasion to do with isolation, loss of power or social standing. We cannot apply such standards of justice to machines we design, and we cannot trace accountability through machines we do not carefully design (Bryson et al., 2017; Bryson and Theodorou, 2019).



## One concerning trend in AI governance is the trend for *value alignment* as a solution to difficult questions of science generally, and AI ethics in particular

Finally, some have expressed concern that it is impossible to maintain regulation in the face of AI because of the rapid rate of change AI entails. It is true that individual humans have limits in their capacities, including on how quickly we can respond. Similarly, legislation can only be written at a particular pace. Latencies are, in fact, deliberately built into the legislative process to ensure the pace of change is not too high for business and personal planning (Holmes, 1988; Cowen, 1992; Ginsburg, 2005; Roithmayr et al., 2015). Therefore legislation alone cannot be expected to keep up with the accelerating pace of change brought on by AI and ICT. I have previously suggested that one mechanism for forging sensible policy is to have domain experts working through professional organizations describe systems of standards (Bryson and Winfield, 2017). The role of government is then reduced to monitoring those efforts and lending enforcement to their outcomes. The arguments I have made above (and in Bryson and Theodorou, 2019) might be seen as a generalization of this principle. Here we are saying that we do not need to change legislation at all, simply hold organizations that build or exploit AI to account for the consequences for their systems' actions by the ordinary and established means of tracing accountability. It is then these organizations who will need to do the innovation on accountability in lock step with their other innovation, so that they can demonstrate that they have always followed due diligence with their systems.

### 5. Conclusion

Artificial intelligence is already changing society at a faster pace than we realize, but at the same time it is not as novel or unique in human experience as we are often led to imagine. Other artifactual entities, such as language and writing, corporations and governments, telecommunication and oil, have previously extended our capacities, altered our economies, and disrupted our social order—generally though not universally for the better. The evidence that we are on average better off for our progress is ironically perhaps the greatest threat we currently need to master: sustainable living and reversing the collapse of biodiversity.

Nevertheless AI—and ICT more generally—may well require radical innovations in the way we govern, and particularly in the way we raise revenue for redistribution. We are faced with transnational wealth transfers through business innovations that have outstripped our capacity to measure or even identify the level of income generated. Further, this new currency

of unknowable value is often personal data, and personal data gives those who hold it the immense power of prediction over the individuals it references.

But beyond the economic and governance challenges, we need to remember that AI first and foremost extends and enhances what it means to be human, and in particular our problem-solving capacities. Given ongoing global challenges such as security and sustainability, such enhancements promise to continue to be of significant benefit, assuming we can establish good mechanisms for their regulation. Through a sensible portfolio of regulatory policies and agencies, we should continue to expand—and also to limit, as appropriate—the scope of potential AI application.



## Acknowledgments

I would like to acknowledge my collaborators, particularly Karine Perset for recruiting me to work on these topics for the OECD and for many good conversations, my (recent) PhD students Andreas Theodorou and Rob Wortham, Alan Winfield with whom some of this content has been reworked and extended to consider the role of professional societies (see Bryson and Winfield, 2017), Karen Croxson of the UK's Financial Conduct Authority, and Will Lowe, particularly for feedback on the sections concerning international relations. Thanks to the OECD for permission to reuse some of the material above for academic contexts such as this volume. I also thank the AXA Research Fund for part-funding my work on AI ethics from 2017–20.

## Notes

1. An older version of some of this material was delivered to the OECD (Karine Perset) in May 2017 under the title “Current and Potential Impacts of Artificial Intelligence and Autonomous Systems on Society,” and contributes to their efforts and documents of late 2018 and early 2019.
2. See further for upside analysis the Obama administration's late AI policy documents (Technology Council Committee on Technology, 2016; of the President, 2016). For reasons of space and focus I also do not discuss here the special case of military use of AI. That is already the subject of other significant academic work, and is generally regulated through different mechanisms than commercial and domestic AI. See, though, Brundage et al. (2018); ICRC (2018).
3. Here, facial recognition.
4. The choice to create life through childbirth is not the same. While we may author some of childrearing, the dispositions just discussed are shared with other primates, and are not options left to parents to author.
5. Importantly, globally, inequality is falling, due to ICT and possibly other progress such as the effective altruism movement and data-lead philanthropy in the developing world. See earlier discussion and Milanovic (2016); Bhorat et al. (2016); Singer (2015); Gabriel (2017).
6. Particularly Nolan McCarty.
7. I personally suspect that some of the advanced political risk-taking, for example in election manipulation, may be a result of those who fear the information age because of its consequences in terms of catching illegal financial conduct, such as money laundering and fraud.
8. This caveat is very important. Much data derived from, for example, governments or commerce may well have strong biases over who is represented or even how well that data is transcribed. Such problems can substantially increase the amount of data required for a given

accuracy of prediction (Meng, 2018).

9. The world is effectively split into two Internets, one inside and one outside the Great Firewall (Ensafi et al., 2015). Both sides similarly contain a small number of extremely successful companies operating in the digital economy (Yiu, 2016; Dolata, 2017).

## Select Bibliography

- Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Green, M., Landau, S., Neumann, P. G., Rivest, R. L., Schiller, J. I., Schneier, B., Specter, M. A., and Weitzner, D. J. 2015. “Keys under doormats: Mandating insecurity by requiring government access to all data and communications.” *Journal of Cybersecurity* 1(1): 69.
- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. 1991. “Complementary DNA sequencing: Expressed sequence tags and human genome project.” *Science* 252(5013): 1651–1656.
- Aker, J. C., and Mbiti, I. M. 2010. “Mobile phones and economic development in Africa.” *Journal of Economic Perspectives* 24(3): 207–232.
- Albrecht, J. P. 2016. “How the GDPR will change the world.” *European Data Protection Law Review* 2(3).
- Armstrong, H., and Read, R. 1995. “Western European micro-states and EU autonomous regions: The advantages of size and sovereignty.” *World Development* 23(7): 1229–1245.
- Assael, Y. M., Shillingford, B., Whiteson, S., and de Freitas, N. 2016. “LipNet: Sentence-level lipreading.” *arXiv preprint arXiv:1611.01599*.
- Author pool. 2017. “I, taxpayer: Why taxing robots is not a good idea—Bill Gates's proposal is revealing about the challenge automation poses.” *The Economist*. print edition.
- Autor, D. H. 2015. “Why are there still so many jobs? The history and future of workplace automation.” *The Journal of Economic Perspectives* 29(3): 3–30.
- Bandyopadhyay, A., and Hazra, A. 2017. “A comparative study of classifier performance on spatial and temporal features of handwritten behavioural data.” In *Intelligent Human Computer Interaction: 8th International Conference, IHCI 2016, Pilani, India, December 12–13, 2016*,

- Proceedings*, A. Basu, S. Das, P. Horain, and S. Bhattacharya (eds.). Cham: Springer International Publishing, 111–121.
- Barber, M. J. 2016. "Ideological donors, contribution limits, and the polarization of American legislatures." *The Journal of Politics* 78(1): 296–310.
- Barnosky, A. D. 2008. "Megafauna biomass tradeoff as a driver of quaternary and future extinctions." *Proceedings of the National Academy of Sciences*, 105(Supplement 1): 11543–11548.
- Barrett, D. P., Barbu, A., Siddharth, N., and Siskind, J. M. 2016. "Saying what you're looking for: Linguistics meets video search." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38(10): 2069–2081.
- Barrows, E. M. 2000. *Animal Behavior Desk Reference: A Dictionary of Animal Behavior, Ecology, and Evolution*. Boca Raton, FL: CRC Press.
- Bhorat, H., Naidoo, K., and Pillay, K. 2016. "Growth, poverty and inequality interactions in Africa: An overview of key issues." UNDP Africa Policy Notes 2016-02, United Nations Development Programme, New York, NY.
- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. London: Springer.
- Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorell, T., Wallis, M., Whitby, B., and Winfield, A. 2011. "Principles of robotics." The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC).
- Bostrom, N. 2012. "The superintelligent will: Motivation and instrumental rationality in advanced artificial agents." *Minds and Machines* 22(2): 71–85.
- Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., and Seaton, D. T. 2013. "Studying learning in the worldwide classroom: Research into edX's first MOOC." *E Research & Practice in Assessment* 8: 13–25.
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hEigeartaigh, S. O., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crotoof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., and Amodei, D. 2018. "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation." Technical report, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, and OpenAI. <https://maliciousaireport.com/>.
- Brundage, M., and Bryson, J. J. 2017. "Smartpolicies for artificial intelligence." In preparation, available as [arXiv:1608.08196](https://arxiv.org/abs/1608.08196).
- Brynjolfsson, E., Rock, D., and Syverson, C. 2017. "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics." In *Economics of Artificial Intelligence*. Chicago: University of Chicago Press.
- Bryson, J. J. 2003. "The behavior-oriented design of modular agent intelligence." In *Agent Technologies, Infrastructures, Tools, and Applications for e-Services*, R. Kowalszyk, J. P. Müller, H. Tianfield, and R. Unland (eds.). Berlin: Springer, 61–76.
- Bryson, J. J. 2008. "Embodiment versus memetics." *Mind & Society* 7(1): 77–94.
- Bryson, J. J. 2015. "Artificial intelligence and pro-social behaviour." In *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, C. Misselhorn (ed.), volume 122 of *Philosophical Studies*. Berlin: Springer, 281–306.
- Bryson, J. J. 2018. "Patience is not a virtue: the design of intelligent systems and systems of ethics." *Ethics and Information Technology* 20(1): 15–26.
- Bryson, J. J., Diamantis, M. E., and Grant, T. D. 2017. "Of, for, and by the people: The legal lacuna of synthetic persons." *Artificial Intelligence and Law* 25(3): 273–291.
- Bryson, J. J., and Theodorou, A. 2019. "How society can maintain human-centric artificial intelligence." In *Human-Centered Digitalization and Services*, M. Toivonen-Noro, and E. Saari (eds.). Springer.
- Bryson, J. J., and Winfield, A. F. T. 2017. "Standardizing ethical design for artificial intelligence and autonomous systems." *Computer* 50(5): 116–119.
- Bullinaria, J. A., and Levy, J. P. 2007. "Extracting semantic representations from word co-occurrence statistics: A computational study." *Behavior Research Methods* 39(3): 510–526.
- Bullough, O. 2018. "The rise of kleptocracy: The dark side of globalization." *Journal of Democracy* 29(1): 25–38.
- Cadwalladr, C. 2017a. "Revealed: How US billionaire helped to back Brexit—Robert Mercer, who bankrolled Donald Trump, played key role with 'sinister' advice on using Facebook data." *The guardian*.
- Cadwalladr, C. 2017b. "Robert Mercer: The big data billionaire waging war on mainstream media." *The guardian*.
- Calinon, S., D'halluin, F., Sauser, E. L., Caldwell, D. G., and Billard, A. G. 2010. "Learning and reproduction of gestures by imitation." *IEEE Robotics & Automation Magazine* 17(2): 44–54.
- Caliskan, A., Bryson, J. J., and Narayanan, A. 2017. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356(6334): 183–186.
- Chessell, M., and Smith, H. C. 2013. *Patterns of Information Management*. London: Pearson Education.
- Christians, A. 2009. "Sovereignty, taxation and social contract." *Minn. J. Int'l L.* 18: 99.
- Chung, J. S., and Zisserman, A. 2017. "Lip reading in the wild." In *Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20–24, 2016, Revised Selected Papers, Part II*, S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato (eds.). Cham: Springer International Publishing, 87–103.
- Claxton, G. 2015. *Intelligence in the Flesh: Why Your Mind Needs Your Body much more than It Thinks*. New Haven: Yale University Press.
- Cooke, M. 1999. "A space of one's own: Autonomy, privacy, liberty." *Philosophy & Social Criticism* 25(1): 22–53.
- Cowen, T. 1992. "Law as a public good: The economics of anarchy." *Economics and Philosophy* 8(02): 249–267.
- Cowen, T. 2011. *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*. Penguin.
- Crabtree, A., and Mortier, R. 2015. "Human data interaction: Historical lessons from social studies and CSCW." In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19–23 September 2015, Oslo, Norway*, N. Boulus-Rødje, G. Ellingsen, T. Bratteteig, M. Anestad, P. and Bjørn (eds.). Cham: Springer International Publishing, 3–21.
- Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Metayer, D. L., Tirtea, R., and Schiffner, S. 2014. "Privacy and data protection by design—from policy to engineering." Technical report, European Union Agency for Network and Information Security (ENISA), Heraklion, Greece.
- Deng, J., Xu, X., Zhang, Z., Frühholz, S., and Schuller, B. 2017. "Universum autoencoder-based domain adaptation for speech emotion recognition." *IEEE Signal Processing Letters* 24(4): 500–504.
- Dennett, D. C. 2013. *Intuition Pumps and Other Tools for Thinking*. New York: WW Norton & Company.
- Dennett, D. C. 2017. *From Bacteria to Bach and Back*. London: Allen Lane.
- Devos, T., and Banaji, M. R. 2003. "Implicit self and identity." *Annals of the New York Academy of Sciences* 1001(1): 177–211.
- Dolata, U. 2017. "Apple, Amazon, Google, Facebook, Microsoft: Market concentration-competition-innovation strategies." SOI Discussion Paper 2017-01, Stuttgarter Beiträge zur Organisations-und Innovationsforschung.



- Editorial Board. 2018. "China's Orwellian tools of high-tech repression." *The Washington Post*.
- Eichhorst, W., and Marx, P. 2011. "Reforming German labour market institutions: A dual path to flexibility." *Journal of European Social Policy* 21(1): 73–87.
- Ensafi, R., Winter, P., Mueen, A., and Crandall, J. R. 2015. "Analyzing the Great Firewall of China over space and time." *Proceedings On Privacy Enhancing Technologies* 2015(1): 61–76.
- Erickson, B. J., Korfiatis, P., Akkus, Z., Kline, T., and Philbrick, K. 2017. "Toolkits and libraries for deep learning." *Journal of Digital Imaging* 30(4): 1–6.
- Eyben, F., Weninger, F., Lehment, N., Schuller, B., and Rigoll, G. 2013. "Affective video retrieval: Violence detection in Hollywood movies by large-scale segmental feature extraction." *PLoS ONE* 8(12): e78506.
- Ford, M. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. London: Oneworld.
- Frischmann, B. M., and Selinger, E. 2016. "Engineering humans with contracts." Research Paper 493, Cardozo Legal Studies. Available at SSRN: <https://ssrn.com/abstract=2834011>.
- Gabriel, I. 2017. "Effective altruism and its critics." *Journal of Applied Philosophy* 34(4): 457–473.
- Gates, C., and Matthews, P. 2014. "Data is the new currency." In *Proceedings of the 2014 New Security Paradigms Workshop*, NSPW '14, New York: ACM, 105–116.
- Gilbert, S. F., Sapp, J., and Tauber, A. I. 2012. "A symbiotic view of life: We have never been individuals." *The Quarterly Review of Biology* 87(4): 325–341. PMID: 23397797.
- Ginsburg, T. 2005. "Locking in democracy: Constitutions, commitment, and international law." *NYUJ Int'l. L. & Pol.* 38: 707.
- Goodman, B., and Flaxman, S. 2016. "EU regulations on algorithmic decision-making and a 'right to explanation.'" In *ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, B. Kim, D. M. Malioutov, K. R. and Varshney (eds.). New York: 26–30.
- Grama, A. 2003. *Introduction To Parallel Computing*. London: Pearson Education.
- Griffin, H. J., Aung, M. S. H., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., and Bianchi-Berthouze, N. 2013. "Laughter type recognition from whole body motion." In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. Geneva: 349–355.
- Haberl, H., Erb, K. H., Krausmann, F., Gaube, V., Bondeau, A., Plutzer, C., Gingrich, S., Lucht, W., and Fischer-Kowalski, M. 2007. "Quantifying and mapping the human appropriation of net primary production in Earth's terrestrial ecosystems." *Proceedings of the National Academy of Sciences* 104(31): 12942–12947.
- Haines, T. S. F., Mac Aodha, O., and Brostow, G. J. 2016. "My text in your handwriting." *ACM Trans. Graph.* 35(3): 26: 1–26: 18.
- Hanahan, D. and Weinberg, R. 2011. "Hallmarks of cancer: The next generation." *Cell* 144(5): 646–674.
- Hancock, J. T., Curry, L. E., Ghorra, S., and Woodworth, M. 2007. "On lying and being lied to: A linguistic analysis of deception in computer-mediated communication." *Discourse Processes* 45(1): 1–23.
- Herrmann, B., Thöni, C., and Gächter, S. 2008. "Antisocial punishment across societies." *Science* 319(5868): 1362–1367.
- Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., and Rigoll, G. 2014. "The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits." *Journal of Visual Communication and Image Representation* 25(1): 195–206.
- Holmes, S. 1988. "Precommitment and the paradox of democracy." *Constitutionalism and Democracy* 195(195): 199–221.
- Hsu, F.-H. 2002. *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*. Princeton, NJ: Princeton University Press.
- Huang, B., Li, M., De Souza, R. L., Bryson, J. J., and Billard, A. 2016. "A modular approach to learning manipulation strategies from human demonstration." *Autonomous Robots* 40(5): 903–927.
- Human Rights Watch. 2018. "Eradicating ideological viruses: China's campaign of repression against Xinjiang's Muslims." Technical report, Human Rights Watch.
- Hunter, L. W., Bernhardt, A., Hughes, K. L., and Skuratowicz, E. 2001. "It's not just the ATMs: Technology, firm strategies, jobs, and earnings in retail banking." *Industrial & Labor Relations Review* 54(2): 402–424.
- ICO. 2018. "Investigation into the use of data analytics in political campaigns: Investigation update." Technical report, Information Commissioner's Office, UK.
- ICRC. 2018. "Ethics and autonomous weapon systems: An ethical basis for human control?" Technical report, International Committee of the Red Cross, Geneva.
- Iyengar, S., Sood, G., and Lelkes, Y. 2012. "Affect, not ideology: Social identity perspective on polarization." *Public Opinion Quarterly* 76(3): 405.
- Jentzsch, N. 2014. "Secondary use of personal data: A welfare analysis." *European Journal of Law and Economics* 44: 1–28.
- Jia, S., Lansdall-Welfare, T., and Cristianini, N. 2016. "Time series analysis of garment distributions via street webcam." In *Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings*, A. Campilho, F. and Karray(eds.). Cham: Springer International Publishing, 765–773.
- Jordan, J. J., Hoffman, M., Nowak, M. A., and Rand, D. G. 2016. "Uncalculating cooperation is used to signal trustworthiness." *Proceedings of the National Academy of Sciences* 113(31): 8658–8663.
- King, G., and Zeng, L. 2001. "Improving forecasts of state failure." *World Politics* 53: 623–658.
- Kleinsmith, A., and Bianchi-Berthouze, N. 2013. "Affective body expression perception and recognition: A survey." *Affective Computing, IEEE Transactions on* 4(1): 15–33.
- Krauss, L. M., and Starkman, G. D. 2000. "Life, the universe, and nothing: Life and death in an ever-expanding universe." *The Astrophysical Journal* 531(1): 22.
- Laland, K. N., Odling-Smee, J., and Feldman, M. W. 2000. "Niche construction, biological evolution, and cultural change." *Behavioral and Brain Sciences* 23(1): 131–146.
- Lamba, S., and Mace, R. 2012. "The evolution of fairness: Explaining variation in bargaining behaviour." *Proceedings of the Royal Society B: Biological Sciences* 280(1750).
- Landau, J.-P. 2016. "Populism and debt: Is Europe different from the U.S.?" Talk at the Princeton Woodrow Wilson School, and in preparation.
- Lawrence, G. W., Kehoe, W. R., Rieger, O. Y., Walters, W. H., and Kenney, A. R. 2000. "Risk management of digital information: A file format investigation." Menlo College Research Paper 93, Council on Library and Information Resources, Washington, D.C.
- Lee, D. S. 1999. "Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage?" *The Quarterly Journal of Economics* 114(3): 977–1023.
- Lee, E., Macvarish, J., and Bristow, J. 2010. "Risk, health and parenting culture." *Health, Risk & Society* 12(4): 293–300.
- Levy, K. E. C. 2015. "Beating the box: Resistance to surveillance in the United States trucking industry." Dissertation chapter and in preparation.
- Liao, S.-H. 2005. "Expert system methodologies and applications: A decade review from 1995 to 2004." *Expert Systems with Applications* 28(1): 93–103.
- List, C. and Pettit, P. 2011. *Group Agency: The Possibility, Design, And Status Of Corporate Agents*. Oxford: Oxford University Press.
- Liu, S., Wang, X., Liu, M., and Zhu, J. 2017. "Towards better analysis of machine learning models: A

- visual analytics perspective." *arXiv preprint arXiv:1702.01226*.
- Lowe, W. 2001. "Towards a theory of semantic space." In *Proceedings of the Twenty-First Annual Meeting of the Cognitive Science Society*, Edinburgh: Lawrence Erlbaum Associates, 576–581.
- Mace, R. 1998. "The co-evolution of human fertility and wealth inheritance strategies." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 353(1367): 389–397.
- Marshall, C. C., Bly, S., and Brun-Cottan, F. 2006. "The long term fate of our digital belongings: Toward a service model for personal archives." *Archiving Conference* 2006(1): 25–30.
- McCarthy, J. 1983. "The little thoughts of thinking machines." *Psychology Today* 17(12): 46–49.
- McCarty, N. M., Poole, K. T., and Rosenthal, H. 2016. *Polarized America: The Dance Of Ideology And Unequal Riches*. Cambridge, MA: MIT Press, 2nd edition.
- Meng, X.-L. 2018. "Statistical paradises and paradoxes in big data (i): Law of large populations, big data paradox, and the 2016 US presidential election." *The Annals of Applied Statistics* 12(2): 685–726.
- Meyer, B. 2016. "Learning to love the government: Trade unions and late adoption of the minimum wage." *World Politics* 68(3): 538–575.
- Miguel, J. C. and Casado, M. Á. 2016. "GAFAnomy (Google, Amazon, Facebook and Apple): The big four and the b-ecosystem." In *Dynamics of Big Internet Industry Groups and Future Trends: A View from Epigenetic Economics*, M. Gómez-Uranga, J. M. Zabala-Iturrigagoitia, J. and Barrutia (eds.). Cham: Springer International Publishing, 127–148.
- Milanovic, B. 2016. *Global inequality*.
- Mill, J. S. 1859. *On Liberty*. London: John W. Parker and Son.
- Miller, T. 2017. *Storming The Wall: Climate Change, Migration, And Homeland Security*. San Francisco: City Lights Books.
- Mishel, L. 2012. "The wedges between productivity and median compensation growth." Issue Brief 330, Washington, DC: Economic Policy Institute.
- Moeslund, T. B., and Granum, E. 2001. "A survey of computer vision-based human motion capture." *Computer Vision and Image Understanding* 81(3): 231–268.
- Morales, A. 2018. "Brexit has already cost the U.K. more than its E.U. budget payments, study shows." *Bloomberg*.
- Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.
- Newman, E. J., Sanson, M., Miller, E. K., Quigley-McBride, A., Foster, J. L., Bernstein, D. M., and Garry, M. 2014. "People with easier to pronounce names promote truthiness of claims." *PLoS ONE* 9(2): e88671.
- Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. 2011. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature." *Decision Support Systems* 50(3): 559–569.
- of the President, E. O. 2016. "Artificial intelligence, automation, and the economy." Technical report, Executive Office of the US President.
- O'Reilly, T. 2017. *WTF? What's the Future and Why It's Up to Us*. New York: Random House.
- Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Person, R. 2018. "Gray zone tactics as asymmetric balancing." Paper presented at the American Political Science Association Annual Meeting.
- Perzanowski, A., and Schultz, J. 2016. *The End of Ownership: Personal Property in the Digital Economy*. Cambridge, MA: MIT Press.
- Pinker, S. 2012. *The Better Angels of Our Nature: The Decline of Violence in History and Its Causes*. London: Penguin.
- Price, G. R. 1972. "Fisher's 'fundamental theorem' made clear." *Annals of Human Genetics* 36(2): 129–140.
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. 2016. "Generative adversarial text to image synthesis." In *Proceedings of the 33rd International Conference on Machine Learning* 48: 1060–1069.
- Roithmayr, D., Isakov, A., and Rand, D. 2015. "Should law keep pace with society? Relative update rates determine the co-evolution of institutional punishment and citizen contributions to public goods." *Games* 6(2): 124.
- Romanes, G. J. 1883. *Animal Intelligence*. New York: D. Appleton.
- Rosner, G. 2014. "Who owns your data?" In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, New York: ACM 623–628.
- Roughgarden, J., Oishi, M., and Akçay, E. 2006. "Reproductive social behavior: Cooperative games to replace sexual selection." *Science* 311(5763): 965–969.
- Russell, S. J., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall, 3rd edition.
- Santoni de Sio, F., and van den Hoven, J. 2018. "Meaningful human control over autonomous systems: A philosophical account." *Frontiers in Robotics and AI* 5: 15.
- Sartori, G., Orru, G., and Monaro, M. 2016. "Detecting deception through kinematic analysis of hand movement." *International Journal of Psychophysiology* 108: 16.
- Scheidt, W. 2017. *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*. Cambridge, MA: Princeton University Press.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. 1996. "Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes." *Proceedings of the National Academy of Sciences* 93(20): 10614–10619.
- Schlager, K. J. 1956. "Systems engineering: Key to modern development." *IRE Transactions on Engineering Management* (3): 64–66.
- Schmitt, J. 2013. "Why does the minimum wage have no discernible effect on employment." Technical Report 22, Center for Economic and Policy Research, Washington, DC.
- Schuller, B., Steidl, S., Batliner, A., Hirschberg, J., Burgoon, J. K., Baird, A., Elkins, A., Zhang, Y., Coutinho, E., and Evanini, K. 2016. "The Interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language." In *Proceedings of Interspeech* 2001–2005.
- Selinger, E. and Hartzog, W. 2017. "Obscurity and privacy." In *Spaces for the Future: A Companion to Philosophy of Technology*, J. Pitt, and A. Shew (eds.). New York: Routledge, in press.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. 2016. "Mastering the game of Go with deep neural networks and tree search." *Nature* 529(7587): 484–489.
- Singer, P. 2015. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*. Melbourne: Text Publishing.
- Singh, J., Pasquier, T., Bacon, J., Ko, H., and Eysers, D. 2016. "Twenty security considerations for cloud-supported Internet of Things." *IEEE Internet of Things Journal* 3(3): 269–284.
- Sipser, M. 2005. *Introduction to the Theory of Computation*. Boston, MA: Thompson, 2nd edition.
- Smith, B. 2018. "Facial recognition technology: The need for public regulation and corporate responsibility." *Microsoft on the Issues*. <https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/>.
- Soares, N., and Fallenstein, B. 2014. "Agent foundations for aligning machine intelligence with human interests: A technical research agenda." Unpublished white paper, available <https://intelligence.org/files/TechnicalAgenda.pdf>.
- Solaiman, S. M. 2017. "Legal personality of robots, corporations, idols and chimpanzees: A quest for legitimacy." *Artificial Intelligence and Law* 25(2): 155–1795.

- Stewart, A. J., McCarty, N., and Bryson, J. J. 2018. “Explaining parochialism: A causal account for political polarization in changing economic environments.” *arXiv preprint arXiv: 1807.11477*.
- Stoddart, D. M. 1990. *The Scented Ape: The Biology and Culture of Human Odour*. Cambridge: Cambridge University Press.
- Suwajanakorn, S., Seitz, S. M., and Kemelmacher-Shlizerman, I. 2017. “Synthesizing Obama: Learning lip sync from audio.” *ACM Transactions on Graphics (TOG)* 36(4): 95.
- Sylwester, K., Mitchell, J., Lowe, W., and Bryson, J. J. 2017. “Punishment as aggression: Uses and consequences of costly punishment across populations.” In preparation.
- Technology Council Committee on Technology, N. S. a. 2016. “Preparing for the future of artificial intelligence.” Technical report, Executive Office of the US President.
- Tepperman, J. 2016. *The Fix: How Countries Use Crises to Solve the World's Worst Problems*. New York: Tim Duggan Books.
- Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., and Nießner, M. 2016. “Face2Face: Real-time face capture and reenactment of RGB videos.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2387–2395.
- Touretzky, D. S. 1988. “On the proper treatment of thermostats.” *Behavioral and Brain Sciences* 11(1): 5556.
- Trewavas, A. 2005. “Green plants as intelligent organisms.” *Trends in Plant Science* 10(9): 413–419.
- Valentino, B. A. 2004. *Final Solutions: Mass Killing and Genocide in the 20th Century*. Ithaca, NY: Cornell University Press.
- Valstar, M. F., and Pantic, M. 2012. “Fully automatic recognition of the temporal phases of facial actions.” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 42(1): 28–43.
- Van Lange, P. A. M., De Bruin, E., Otten, W., and Joireman, J. A. 1997. “Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence.” *Journal of Personality and Social Psychology* 73(4): 733–746.
- Van Schaik, C., Graber, S., Schuppli, C., and Burkart, J. 2017. “The ecology of social learning in animals and its link with intelligence.” *The Spanish Journal of Psychology* 19.
- Vincent, J. 2016. “Artificial intelligence is going to make it easier than ever to fake images and video.” *The Verge*.
- Weber, R. H. 2010. “Internet of Things: New security and privacy challenges.” *Computer Law & Security Review* 26(1): 23–30.
- Widrow, B., Rumelhart, D. E., and Lehr, M. A. 1994. “Neural networks: Applications in industry, business and science.” *Commun. ACM* 37(3): 93–105.
- Williams, C. P. 2010. *Explorations in Quantum Computing*. London: Springer-Verlag.
- Winston, P. H. 1984. *Artificial Intelligence*. Boston, MA: Addison-Wesley.
- Wolpert, D. H. 1996a. “The existence of a priori distinctions between learning algorithms.” *Neural Computation* 8(7): 1391–1420.
- Wolpert, D. H. 1996b. “The lack of a priori distinctions between learning algorithms.” *Neural Computation* 8(7): 1341–1390.
- Wright, G. 1974. “The political economy of new deal spending: An econometric analysis.” *The Review of Economics and Statistics* 56(1): 30–38.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., ukasz Kaiser, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. 2016. “Google’s neural machine translation system: Bridging the gap between human and machine translation.” *CoRR*, abs/1609.08144.
- Yan, X. 2006. “The rise of China and its power status.” *The Chinese Journal of International Politics* 1(1): 5–33.
- Yiu, E. 2016. “Alibaba tops Tencent as Asia’s biggest company by market value: New-economy companies that owe their revenue to technology and the Internet are now more valuable than oil refineries, manufacturers and banks.” *South China Morning Post*.
- Youyou, W., Kosinski, M., and Stillwell, D. 2015. “Computer-based personality judgments are more accurate than those made by humans.” *Proceedings of the National Academy of Sciences* 112(4): 1036–1040.



**Ramón López de Mántaras**  
Institute of the Spanish  
National Research Council

Ramón López de Mántaras is a Research Professor of the Spanish National Research Council (CSIC) and Director of the Artificial Intelligence Research Institute (IIIA). He has an MSc in Computer Science from the University of California Berkeley, a PhD in Physics (Automatic Control) from the University of Toulouse, and a PhD in Computer Science from the Technical University of Barcelona. López de Mántaras is a pioneer of artificial intelligence (AI) in Spain, with contributions, since 1976, in pattern recognition, approximate reasoning, expert systems, machine learning, case-based reasoning, autonomous robots, and AI and music. He is the author of nearly 300 papers and an invited plenary speaker at numerous international conferences. Former Editor-in-Chief of *AI Communications*, he is an editorial board member of several top international journals, an ECCAI Fellow, and the co-recipient of five best paper awards at international conferences. Among other awards, he has received the "City of Barcelona" Research Prize in 1981, the 2011 "American Association of Artificial Intelligence (AAAI) Robert S. Englemore Memorial Award," the 2012 "Spanish National Computer Science Award" from the Spanish Computer Society, the 2016 "Distinguished Service Award" of the European Association of Artificial Intelligence (EurAI), and the 2017 "IJCAI Donald E. Walker Distinguished Service Award." He is a member of the Institut d'Estudis Catalans (IEC). He serves on a variety of panels and advisory committees for public and private institutions based in the USA and Europe, such as the EC Joint Research Center High-Level Peer Group. He is presently working on case-based reasoning, machine learning, and AI applications to music.

Recommended books: *The Master Algorithm*, Pedro Domingos, Basic Books, 2015. *Inteligencia Artificial*, Ramón López de Mántaras and Pedro Meseguer, Los Libros de la Catarata, 2017.

**This article contains some reflections about artificial intelligence (AI). First, the distinction between strong and weak AI and the related concepts of general and specific AI is made, making it clear that all existing manifestations of AI are weak and specific. The main existing models are briefly described, insisting on the importance of corporality as a key aspect to achieve AI of a general nature. Also discussed is the need to provide common-sense knowledge to the machines in order to move toward the ambitious goal of building general AI. The paper also looks at recent trends in AI based on the analysis of large amounts of data that have made it possible to achieve spectacular progress very recently, also mentioning the current difficulties of this approach to AI. The final part of the article discusses other issues that are and will continue to be vital in AI and closes with a brief reflection on the risks of AI.**



## Introduction



The final goal of artificial intelligence (AI)—that a machine can have a type of *general* intelligence similar to a human's—is one of the most ambitious ever proposed by science. In terms of difficulty, it is comparable to other great scientific goals, such as explaining the origin of life or the Universe, or discovering the structure of matter. In recent centuries, this interest in building intelligent machines has led to the invention of models or metaphors of the human brain. In the seventeenth century, for example, Descartes wondered whether a complex mechanical system of gears, pulleys, and tubes could possibly emulate thought. Two centuries later, the metaphor had become telephone systems, as it seemed possible that their connections could be likened to a neural network. Today, the dominant model is computational and is based on the digital computer. Therefore, that is the model we will address in the present article.

### The Physical Symbol System Hypothesis: Weak AI Versus Strong AI

In a lecture that coincided with their reception of the prestigious Turing Prize in 1975, Allen Newell and Herbert Simon (Newell and Simon, 1976) formulated the “Physical Symbol System” hypothesis, according to which “a physical symbol system has the necessary and sufficient means for general intelligent action.” In that sense, given that human beings are able to display intelligent behavior in a general way, we, too, would be physical symbol systems. Let us clarify what Newell and Simon mean when they refer to a Physical Symbol System (PSS). A PSS consists of a set of entities called symbols that, through relations, can be combined to form larger structures—just as atoms combine to form molecules—and can be transformed by applying a set of processes. Those processes can create new symbols, create or modify relations among symbols, store symbols, detect whether two are the same or different, and so on. These symbols are physical in the sense that they have an underlying physical-electronic layer (in the case of computers) or a physical-biological one (in the case of human beings). In fact, in the case of computers, symbols are established through digital electronic circuits, whereas humans do so with neural networks. So, according to the PSS hypothesis, the nature of the underlying layer (electronic circuits or neural networks) is unimportant as long as it allows symbols to be processed. Keep in mind that this is a hypothesis, and should, therefore, be neither accepted nor rejected a priori. Either way, its validity or refutation must be verified according to the scientific method, with experimental testing. AI is precisely the scientific field dedicated to attempts to verify this hypothesis in the context of digital computers, that is, verifying whether a properly programmed computer is capable of general intelligent behavior.

Specifying that this must be general intelligence rather than specific intelligence is important, as human intelligence is also general. It is quite a different matter to exhibit specific intelligence. For example, computer programs capable of playing chess at Grand-Master levels are incapable of playing checkers, which is actually a much simpler game. In order for the same computer to play checkers, a different, independent program must be designed and executed. In other words, the computer cannot draw on its capacity to play chess as a means of adapting to the game of checkers. This is not the case, however, with humans, as any human chess player can take advantage of his knowledge of that game to play checkers perfectly in a matter of minutes. The design and application of artificial intelligences that can only behave intelligently in a very specific setting is related to what is known as *weak AI*, as opposed to *strong AI*. Newell, Simon, and the other founding fathers of AI refer to the



latter. Strictly speaking, the PSS hypothesis was formulated in 1975, but, in fact, it was implicit in the thinking of AI pioneers in the 1950s and even in Alan Turing's groundbreaking texts (Turing, 1948, 1950) on intelligent machines.

This distinction between weak and strong AI was first introduced by philosopher John Searle in an article criticizing AI in 1980 (Searle, 1980), which provoked considerable discussion at the time, and still does today. Strong AI would imply that a properly designed computer does not simulate a mind but *actually is one*, and should, therefore, be capable of an intelligence equal, or even superior to human beings. In his article, Searle sought to demonstrate that strong AI is impossible, and, at this point, we should clarify that general AI is not the same as strong AI. Obviously they are connected, but only in one sense: all strong AI will necessarily be general, but there can be general AIs capable of multitasking but not strong in the sense that, while they can emulate the capacity to exhibit general intelligence similar to humans, they do not experience states of mind.

**The final goal of AI—that a machine can have a type of general intelligence similar to a human's—is one of the most ambitious ever proposed by science. In terms of difficulty, it is comparable to other great scientific goals, such as explaining the origin of life or the Universe, or discovering the structure of matter**

According to Searle, weak AI would involve constructing programs to carry out specific tasks, obviously without need for states of mind. Computers' capacity to carry out specific tasks, sometimes even better than humans, has been amply demonstrated. In certain areas, weak AI has become so advanced that it far outstrips human skill. Examples include solving logical formulas with many variables, playing chess or Go, medical diagnosis, and many others relating to decision-making. Weak AI is also associated with the formulation and testing of hypotheses about aspects of the mind (for example, the capacity for deductive reasoning, inductive learning, and so on) through the construction of programs that carry out those functions, even when they do so using processes totally unlike those of the human brain. As of today, absolutely all advances in the field of AI are manifestations of weak and specific AI.

#### **The Principal Artificial Intelligence Models: Symbolic, Connectionist, Evolutionary, and Corporeal**

The *symbolic* model that has dominated AI is rooted in the PSS model and, while it continues to be very important, is now considered classic (it is also known as GOFAI, that is, *Good Old-Fashioned AI*). This top-down model is based on logical reasoning and heuristic searching as the pillars of problem solving. It does not call for an intelligent system to be part of a body, or to be situated in a real setting. In other words, symbolic AI works with abstract representations of the real world that are modeled with representational languages based primarily on mathematical logic and its extensions. That is why the first intelligent systems mainly solved problems that did not require direct interaction with the environment, such as demonstrating simple mathematical theorems or playing chess—in fact, chess programs need neither visual perception for seeing the board, nor technology to actually move

the pieces. That does not mean that symbolic AI cannot be used, for example, to program the reasoning module of a physical robot situated in a real environment, but, during its first years, AI's pioneers had neither languages for representing knowledge nor programming that could do so efficiently. That is why the early intelligent systems were limited to solving problems that did not require direct interaction with the real world. Symbolic AI is still used today to demonstrate theorems and to play chess, but it is also a part of applications that require perceiving the environment and acting upon it, for example learning and decision-making in autonomous robots.

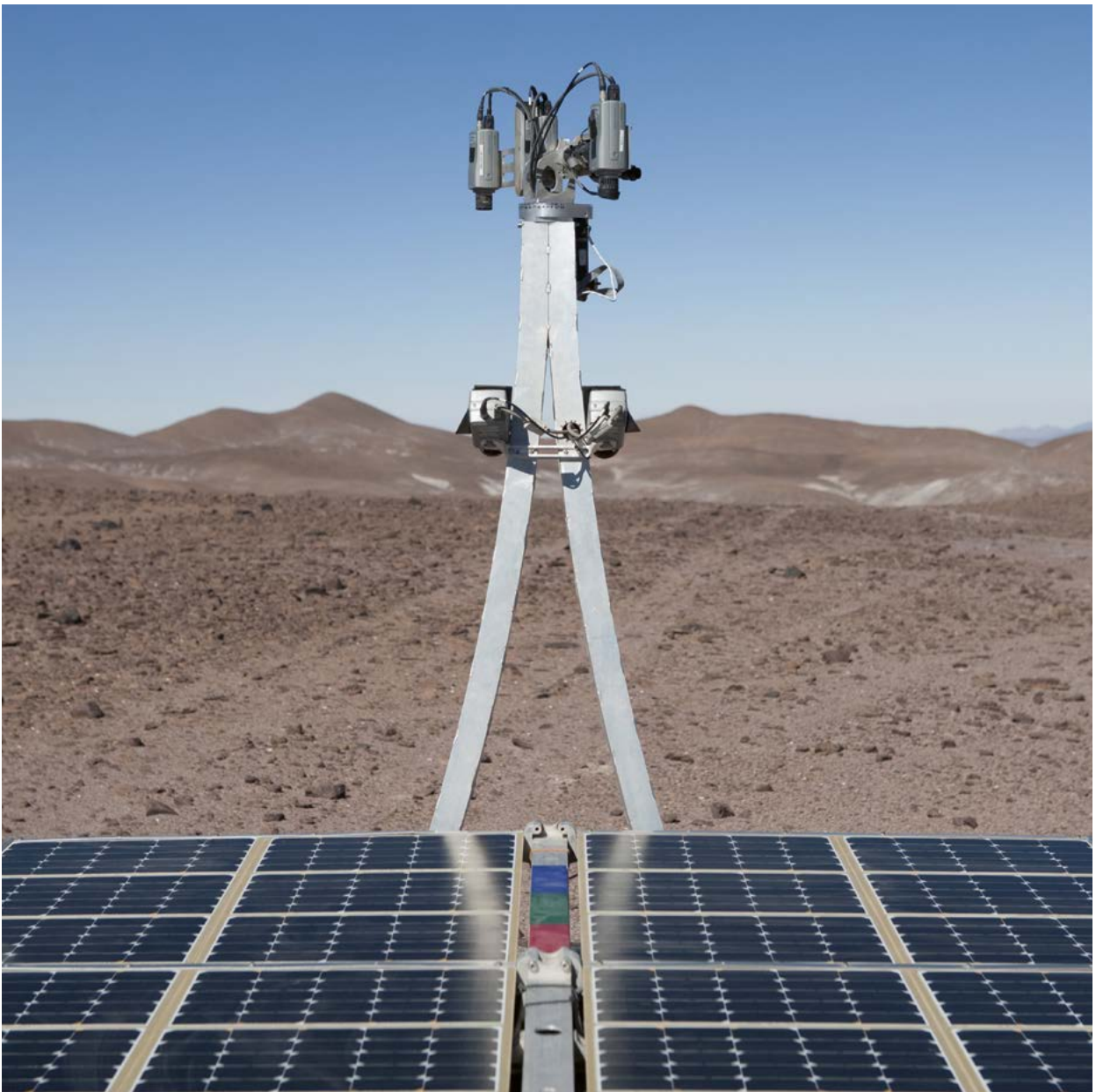


**The symbolic model that has dominated AI is rooted in the PSS model and, while it continues to be very important, is now considered classic (it is also known as GOFAI, that is, Good Old-Fashioned AI). This top-down model is based on logical reasoning and heuristic searching as the pillars of problem solving**

At the same time that symbolic AI was being developed, a biologically based approach called *connectionist* AI arose. Connectionist systems are not incompatible with the PSS hypothesis but, unlike symbolic AI, they are modeled from the bottom up, as their underlying hypothesis is that intelligence emerges from the distributed activity of a large number of interconnected units whose models closely resemble the electrical activity of biological neurons. In 1943, McCulloch and Pitts (1943) proposed a simplified model of the neuron based in the idea that it is essentially a logic unit. This model is a mathematical abstraction with inputs (dendrites) and outputs (axons). The output value is calculated according to the result of a weighted sum of the entries in such a way that if that sum surpasses a preestablished threshold, it functions as a “1,” otherwise it will be considered a “0.” Connecting the output of each neuron to the inputs of other neurons creates an artificial neural network. Based on what was then known about the reinforcement of synapses among biological neurons, scientists found that these artificial neural networks could be trained to learn functions that related inputs to outputs by adjusting the weights used to determine connections between neurons. These models were hence considered more conducive to learning, cognition, and memory than those based on symbolic AI. Nonetheless, like their symbolic counterparts, intelligent systems based on connectionism do not need to be part of a body, or situated in real surroundings. In that sense, they have the same limitations as symbolic systems. Moreover, real neurons have complex dendritic branching with truly significant electrical and chemical properties. They can contain ionic conductance that produces nonlinear effects. They can receive tens of thousands of synapses with varied positions, polarities, and magnitudes. Furthermore, most brain cells are not neurons, but rather *glial* cells that not only regulate neural functions but also possess electrical potentials, generate calcium waves, and communicate with others. This would seem to indicate that they play a very important role in cognitive processes, but no existing connectionist models include glial cells so they are, at best, extremely incomplete and, at worst, erroneous. In short, the enormous complexity of the brain is very far indeed from current models. And that very complexity also raises the idea of what has come to be known as *singularity*, that is, future artificial superintelligences based on replicas of the brain



Engineers at Carnegie Mellon University developed this robot named Zoe to detect life in apparently uninhabited environments. Zoe includes a cutting-edge system for detecting organic molecules which may help to find life on Mars. It is twenty times faster than the Mars explorer robots, Spirit and Opportunity. Atacama Desert, Chile, 2005







00

04:58

nico

電王戦





**The design and application of artificial intelligences that can only behave intelligently in a very specific setting is related to what is known as weak AI, as opposed to strong AI**

Masayuki Toyoshima, a professional shogi or Japanese chess player, plays against a YSS computer program that moves the pieces with a robotic arm. Osaka, March, 2014

but capable, in the coming twenty-five years, of far surpassing human intelligence. Such predictions have little scientific merit.

Another biologically inspired but non-corporeal model that is also compatible with the PSS hypothesis is *evolutionary computation* (Holland, 1975). Biology's success at evolving complex organisms led some researchers from the early 1960s to consider the possibility of imitating evolution. Specifically, they wanted computer programs that could evolve, automatically improving solutions to the problems for which they had been programmed. The idea being that, thanks to mutation operators and crossed "chromosomes" modeled by those programs, they would produce new generations of modified programs whose solutions would be better than those offered by the previous ones. Since we can define AI's goal as the search for programs capable of producing intelligent behavior, researchers thought that evolutionary programming might be used to find those programs among all possible programs. The reality is much more complex, and this approach has many limitations although it has produced excellent results in the resolution of optimization problems.



## **The human brain is very far removed indeed from AI models, which suggests that so-called singularity—artificial superintelligences based on replicas of the brain that far surpass human intelligence—are a prediction with very little scientific merit**

One of the strongest critiques of these non-corporeal models is based on the idea that an intelligent agent needs a body in order to have direct experiences of its surroundings (we would say that the agent is "situated" in its surroundings) rather than working from a programmer's abstract descriptions of those surroundings, codified in a language for representing that knowledge. Without a body, those abstract representations have no semantic content for the machine, whereas direct interaction with its surroundings allows the agent to relate signals perceived by its sensors to symbolic representations generated on the basis of what has been perceived. Some AI experts, particularly Rodney Brooks (1991), went so far as to affirm that it was not even necessary to generate those internal representations, that is, that an agent does not even need an internal representation of the world around it because the world itself is the best possible model of itself, and most intelligent behavior does not require reasoning, as it emerged directly from interaction between the agent and its surroundings. This idea generated considerable argument, and some years later, Brooks himself admitted that there are many situations in which an agent requires an internal representation of the world in order to make rational decisions.

In 1965, philosopher Hubert Dreyfus affirmed that AI's ultimate objective—strong AI of a general kind—was as unattainable as the seventeenth-century alchemists' goal of transforming lead into gold (Dreyfus, 1965). Dreyfus argued that the brain processes information in a global and continuous manner, while a computer uses a finite and discreet set of deterministic operations, that is, it applies rules to a finite body of data. In that sense, his argument resembles Searle's, but in later articles and books (Dreyfus, 1992), Dreyfus argued that the body plays a crucial role in intelligence. He was thus one of the first to advocate the need for intelligence to be part of a body that would allow it to interact with the world. The main idea is that living beings' intelligence derives from their situation in surroundings with which they



can interact through their bodies. In fact, this need for corporeality is based on Heidegger's phenomenology and its emphasis on the importance of the body, its needs, desires, pleasures, suffering, ways of moving and acting, and so on. According to Dreyfus, AI must model all of those aspects if it is to reach its ultimate objective of strong AI. So Dreyfus does not completely rule out the possibility of strong AI, but he does state that it is not possible with the classic methods of symbolic, non-corporeal AI. In other words, he considers the Physical Symbol System hypothesis incorrect. This is undoubtedly an interesting idea and today it is shared by many AI researchers. As a result, the corporeal approach with internal representation has been gaining ground in AI and many now consider it essential for advancing toward general intelligences. In fact, we base much of our intelligence on our sensory and motor capacities. That is, the body shapes intelligence and therefore, without a body general intelligence cannot exist. This is so because the body as hardware, especially the mechanisms of the sensory and motor systems, determines the type of interactions that an agent can carry out. At the same time, those interactions shape the agent's cognitive abilities, leading to what is known as *situated cognition*. In other words, as occurs with human beings, the machine is situated in real surroundings so that it can have interactive experiences that will eventually allow it to carry out something similar to what is proposed in Piaget's cognitive development theory (Inhelder and Piaget, 1958): a human being follows a process of mental maturity in stages and the different steps in this process may possibly work as a guide for designing intelligent machines. These ideas have led to a new sub-area of AI called *development robotics* (Weng et al., 2001).

### Specialized AI's Successes

All of AI's research efforts have focused on constructing specialized artificial intelligences, and the results have been spectacular, especially over the last decade. This is thanks to the combination of two elements: the availability of huge amounts of data, and access to high-level computation for analyzing it. In fact, the success of systems such as AlphaGO (Silver et al., 2016), Watson (Ferrucci et al., 2013), and advances in autonomous vehicles or image-based medical diagnosis have been possible thanks to this capacity to analyze huge amounts of data and efficiently detect patterns. On the other hand, we have hardly advanced at all in the quest for general AI. In fact, we can affirm that current AI systems are examples of what Daniel Dennet called "competence without comprehension" (Dennet, 2018).

**All of AI's research efforts have focused on constructing specialized artificial intelligences, and the results have been spectacular, especially over the last decade. This is thanks to the combination of two elements: the availability of huge amounts of data, and access to high-level computation for analyzing it**

Perhaps the most important lesson we have learned over the last sixty years of AI is that what seemed most difficult (diagnosing illnesses, playing chess or Go at the highest level) have turned out to be relatively easy, while what seemed easiest has turned out to be the most





Porior millenem exceaue corat et rempore officatemqui  
con nonsedit aut que repel et eic to iust, consequid  
quundis doluptur, ullat hicilitio eum recte est ut aut lab  
id ex et dolupta tioria deni re, oditis invero nsent, susam  
remoles diaestem voloreh endaect inciam, conse pro





difficult of all. The explanation of this apparent contradiction may be found in the difficulty of equipping machines with the knowledge that constitutes “common sense.” without that knowledge, among other limitations, it is impossible to obtain a deep understanding of language or a profound interpretation of what a visual perception system captures. Common-sense knowledge is the result of our lived experiences. Examples include: “water always flows downward;” “to drag an object tied to a string, you have to pull on the string, not push it;” “a glass can be stored in a cupboard, but a cupboard cannot be stored in a glass;” and so on. Humans easily handle millions of such common-sense data that allow us to understand the world we inhabit. A possible line of research that might generate interesting results about the acquisition of common-sense knowledge is the development robotics mentioned above. Another interesting area explores the mathematical modeling and learning of cause-and-effect relations, that is, the learning of causal, and thus asymmetrical, models of the world. Current systems based on deep learning are capable of learning symmetrical mathematical functions, but unable to learn asymmetrical relations. They are, therefore, unable to distinguish cause from effects, such as the idea that the rising sun causes a rooster to crow, but not vice versa (Pearl and Mackenzie, 2018; Lake et al., 2016).

### The Future: Toward Truly Intelligent Artificial Intelligences

The most complicated capacities to achieve are those that require interacting with unrestricted and not previously prepared surroundings. Designing systems with these capabilities requires the integration of development in many areas of AI. We particularly need knowledge-representation languages that codify information about many different types of objects, situations, actions, and so on, as well as about their properties and the relations among them—especially, cause-and-effect relations. We also need new algorithms that can use these representations in a robust and efficient manner to resolve problems and answer questions on almost any subject. Finally, given that they will need to acquire an almost unlimited amount of knowledge, those systems will have to be able to learn continuously throughout their existence. In sum, it is essential to design systems that combine perception, representation, reasoning, action, and learning. This is a very important AI problem as we still do not know how to integrate all of these components of intelligence. We need cognitive architectures (Forbus, 2012) that integrate these components adequately. Integrated systems are a fundamental first step in someday achieving general AI.

**The most complicated capacities to achieve are those that require interacting with unrestricted and not previously prepared surroundings. Designing systems with these capabilities requires the integration of development in many areas of AI**

Among future activities, we believe that the most important research areas will be hybrid systems that combine the advantages of systems capable of reasoning on the basis of knowledge and memory use (Graves et al., 2016) with those of AI based on the analysis of massive amounts of data, that is, deep learning (Bengio, 2009). Today, deep-learning systems are significantly limited by what is known as “catastrophic forgetting.” This means that if they



have been trained to carry out one task (playing Go, for example) and are then trained to do something different (distinguishing between images of dogs and cats, for example) they completely forget what they learned for the previous task (in this case, playing Go). This limitation is powerful proof that those systems do not learn anything, at least in the human sense of learning. Another important limitation of these systems is that they are “black boxes” with no capacity to explain. It would, therefore, be interesting to research how to endow deep-learning systems with an explicative capacity by adding modules that allow them to explain how they reached the proposed results and conclusion, as the capacity to explain is an essential characteristic of any intelligent system. It is also necessary to develop new learning algorithms that do not require enormous amounts of data to be trained, as well as much more energy-efficient hardware to implement them, as energy consumption could end up being one of the main barriers to AI development. Comparatively, the brain is various orders of magnitude more efficient than the hardware currently necessary to implement the most sophisticated AI algorithms. One possible path to explore is memristor-based neuromorphic computing (Saxena et al., 2018).

Other more classic AI techniques that will continue to be extensively researched are multi-agent systems, action planning, experience-based reasoning, artificial vision, multimodal person-machine communication, humanoid robotics, and particularly, new trends in *development robotics*, which may provide the key to endowing machines with common sense, especially the capacity to learn the relations between their actions and the effects these produce on their surroundings. We will also see significant progress in biomimetic approaches to reproducing animal behavior in machines. This is not simply a matter of reproducing an animal’s behavior, it also involves understanding how the brain that produces that behavior actually works. This involves building and programming electronic circuits that reproduce the cerebral activity responsible for this behavior. Some biologists are interested in efforts to create the most complex possible artificial brain because they consider it a means of better understanding that organ. In that context, engineers are seeking biological information that makes designs more efficient. Molecular biology and recent advances in optogenetics will make it possible to identify which genes and neurons play key roles in different cognitive activities.

## **Development robotics may provide the key to endowing machines with common sense, especially the capacity to learn the relations between their actions and the effects these produce on their surroundings**

As to applications: some of the most important will continue to be those related to the Web, video-games, personal assistants, and autonomous robots (especially autonomous vehicles, social robots, robots for planetary exploration, and so on). Environmental and energy-saving applications will also be important, as well as those designed for economics and sociology. Finally, AI applications for the arts (visual arts, music, dance, narrative) will lead to important changes in the nature of the creative process. Today, computers are no longer simply aids to creation; they have begun to be creative agents themselves. This has led to a new and very promising AI field known as *computational creativity* which is producing very interesting results (Colton et al., 2009, 2015; López de Mántaras, 2016) in chess, music, the visual arts, and narrative, among other creative activities.



No matter how intelligent future artificial intelligences become—even general ones—they will never be the same as human intelligences. As we have argued, the mental development needed for all complex intelligence depends on interactions with the environment and those interactions depend, in turn, on the body—especially the perceptive and motor systems. This, along with the fact that machines will not follow the same socialization and culture-acquisition processes as ours, further reinforces the conclusion that, no matter how sophisticated they become, these intelligences will be different from ours. The existence of intelligences unlike ours, and therefore alien to our values and human needs, calls for reflection on the possible ethical limitations of developing AI. Specifically, we agree with Weizenbaum's affirmation (Weizenbaum, 1976) that no machine should ever make entirely autonomous decisions or give advice that call for, among other things, wisdom born of human experiences, and the recognition of human values.

**No matter how intelligent future artificial intelligences become, they will never be the same as human intelligence: the mental development needed for all complex intelligence depends on interactions with the environment and those interactions depend, in turn, on the body—especially the perceptive and motor systems**

The true danger of AI is not the highly improbable technological singularity produced by the existence of hypothetical future artificial superintelligences; the true dangers are already here. Today, the algorithms driving Internet search engines or the recommendation and personal-assistant systems on our cellphones, already have quite adequate knowledge of what we do, our preferences and tastes. They can even infer what we think about and how we feel. Access to massive amounts of data that we generate voluntarily is fundamental for this, as the analysis of such data from a variety of sources reveals relations and patterns that could not be detected without AI techniques. The result is an alarming loss of privacy. To avoid this, we should have the right to own a copy of all the personal data we generate, to control its use, and to decide who will have access to it and under what conditions, rather than it being in the hands of large corporations without knowing what they are really doing with our data.

AI is based on complex programming, and that means there will inevitably be errors. But even if it were possible to develop absolutely dependable software, there are ethical dilemmas that software developers need to keep in mind when designing it. For example, an autonomous vehicle could decide to run over a pedestrian in order to avoid a collision that could harm its occupants. Outfitting companies with advanced AI systems that make management and production more efficient will require fewer human employees and thus generate more unemployment. These ethical dilemmas are leading many AI experts to point out the need to regulate its development. In some cases, its use should even be prohibited. One clear example is autonomous weapons. The three basic principles that govern armed conflict: discrimination (the need to distinguish between combatants and civilians, or between





a combatant who is surrendering and one who is preparing to attack), proportionality (avoiding the disproportionate use of force), and precaution (minimizing the number of victims and material damage) are extraordinarily difficult to evaluate and it is therefore almost impossible for the AI systems in autonomous weapons to obey them. But even if, in the very long term, machines were to attain this capacity, it would be indecent to delegate the decision to kill to a machine. Beyond this kind of regulation, it is imperative to educate the citizenry as to the risks of intelligent technologies, and to insure that they have the necessary competence for controlling them, rather than being controlled *by* them. Our future citizens need to be much more informed, with a greater capacity to evaluate technological risks, with a greater critical sense and a willingness to exercise their rights. This training process must begin at school and continue at a university level. It is particularly necessary for science and engineering students to receive training in ethics that will allow them to better grasp the social implications of the technologies they will very likely be developing. Only when we invest in education will we achieve a society that can enjoy the advantages of intelligent technology while minimizing the risks. AI unquestionably has extraordinary potential to benefit society, as long as we use it properly and prudently. It is necessary to increase awareness of AI's limitations, as well as to act collectively to guarantee that AI is used for the common good, in a safe, dependable, and responsible manner.

The road to truly intelligent AI will continue to be long and difficult. After all, this field is barely sixty years old, and, as Carl Sagan would have observed, sixty years are barely the blink of an eye on a cosmic time scale. Gabriel García Márquez put it more poetically in a 1936 speech ("The Cataclysm of Damocles"): "Since the appearance of visible life on Earth, 380 million years had to elapse in order for a butterfly to learn how to fly, 180 million years to create a rose with no other commitment than to be beautiful, and four geological eras in order for us human beings to be able to sing better than birds, and to be able to die from love."

## Select Bibliography

- Bengio, Y. 2009. "Learning deep architectures for AI." *Foundations and Trends in Machine Learning* 2(1): 1–127.
- Brooks, R. A. 1991. "Intelligence without reason." *IJCAI-91 Proceedings of the Twelfth International Joint Conference on Artificial Intelligence* 1: 569–595.
- Colton, S., Lopez de Mantaras, R., and Stock, O. 2009. "Computational creativity: Coming of age." *AI Magazine* 30(3): 11–14.
- Colton, S., Halskov, J., Ventura, D., Gouldstone, I., Cook, M., and Pérez-Ferrer, B. 2015. "The Painting Fool sees! New projects with the automated painter." *International Conference on Computational Creativity (ICCC 2015)*: 189–196.
- Dennet, D. C. 2018. *From Bacteria to Bach and Back: The Evolution of Minds*. London: Penguin.
- Dreyfus, H. 1965. *Alchemy and Artificial Intelligence*. Santa Monica: Rand Corporation.
- Dreyfus, H. 1992. *What Computers Still Can't Do*. New York: MIT Press.
- Ferrucci, D. A., Levas, A., Bagchi, S., Gondek, D., and Mueller, E. T. 2013. "Watson: Beyond jeopardy!" *Artificial Intelligence* 199: 93–105.
- Forbus, K. D. 2012. "How minds will be built." *Advances in Cognitive Systems* 1: 47–58.
- Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Gómez-Colmenarejo, S., Grefenstette, E., Ramalho, T., Agapiou, J., Puigdomènech-Badia, A., Hermann, K. M., Zwols, Y., Ostrovski, G., Cain, A., King, H., Summerfield, C., Blunsom, P., Kavukcuoglu, K., and Hassabis, D. 2016. "Hybrid computing using a neural network with dynamic external memory." *Nature* 538: 471–476.
- Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Inhelder, B., and Piaget, J. 1958. *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. 2017. "Building machines that learn and think like people." *Behavioral and Brain Sciences* 40:e253.
- López de Mantaras, R. 2016. "Artificial intelligence and the arts: Toward computational creativity." In *The Next Step: Exponential Life*. Madrid: BBVA Open Mind, 100–125.
- McCulloch, W. S., and Pitts, W. 1943. "A logical calculus of ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics* 5: 115–133.
- Newell, A., and Simon, H. A. 1976. "Computer science as empirical inquiry: Symbols and search." *Communications of the ACM* 19(3): 113–126.
- Pearl, J., and Mackenzie, D. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.
- Saxena, V., Wu, X., Srivastava, I., and Zhu, K. 2018. "Towards neuromorphic learning machines using emerging memory devices with brain-like energy efficiency." Preprints: <https://www.preprints.org/manuscript/201807.0362/v1>.
- Searle, J. R. 1980. "Minds, brains, and programs," *Behavioral and Brain Sciences* 3(3): 417–457.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., ven den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. 2016. "Mastering the game of Go with deep neural networks and tree search." *Nature* 529(7587): 484–489.
- Turing, A. M. 1948. "Intelligent machinery." National Physical Laboratory Report. Reprinted in: *Machine Intelligence* 5, B. Meltzer and D. Michie (eds.). Edinburgh: Edinburgh University Press, 1969.
- Turing, A. M. 1950. "Computing machinery and intelligence." *Mind* LIX(236): 433–460.
- Weizenbaum, J. 1976. *Computer Power and Human Reasoning: From Judgment to Calculation*. San Francisco: W. H. Freeman and Co.
- Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. 2001. "Autonomous mental development by robots and animals." *Science* 291: 599–600.



**José M. Mato**

CIC bioGUNE. Center for Cooperative Research in Biosciences & CIC biomaGUNE. Center for Cooperative Research in Biomaterials

José M. Mato is the founder and General Director of the Research Centers bioGUNE (Bilbao) and biomaGUNE (San Sebastián), and Research Professor of the Spanish National Research Council (CSIC), Spain. A graduate in Biochemistry at the University of Madrid, Professor Mato received a PhD from Leiden University and was awarded the C. J. Kok prize for his thesis. He was a postdoctoral fellow at the Biozentrum of the University of Basel and the National Institutes of Health, and a faculty member at the Jiménez Díaz Foundation in Madrid before being named Research Professor at the CSIC. He has been Professor of the Faculty of Medicine of the University of Navarra and Visiting Professor at the University of Pennsylvania and Thomas Jefferson University. From 1992 to 1996, Professor Mato was President of the CSIC and in 2004 was awarded the Spanish National Research Prize in Medicine. The aim of his work is to study metabolic alterations as a tool and target for the detection, prevention, and treatment of nonalcoholic steatohepatitis, including its progression to liver cirrhosis and cancer. He is the cofounder of OWL Metabolomics in 2000 and of ATLAS Molecular Pharma in 2016.

Recommended book: *The Discoverers*, Daniel J. Boorstin, Random House, 1983.

**An increase of three decades in the life expectancy of persons born in numerous countries and regions of the West during the twentieth century is the clearest exponent of the social value of science and the triumph of public-health policies over illness. Continuing to pursue this mission, expanding it to cover the thousands of rare diseases still untreated, and to cover all countries and all of their citizens is our challenge for the coming decades.**



There are five viewpoints for any illness: that of the patient, the clinic, the scientist, industry, and the regulatory agencies. Initially, the patient experiences illness and medical attention in a single, individual way, but that individuality later expands to encompass the community of people who share the same symptoms. This connection between the individual and his or her illness has led to what the World Health Organization calls patient empowerment, “a process through which people gain greater control over decisions and actions affecting their health,”<sup>1</sup> which must, in turn, be viewed as both an individual and a collective process.

The clinic, in turn, does not view each patient as a unique and individual case; it has learned to avoid emotional ties with its patients. They gradually stop having names and turn into illnesses that require precise diagnosis and adequate treatment.

Some augur that technology will replace a high percentage of doctors in the future. I believe that a considerable part of the tasks currently being carried out will someday be handled by robots, that teams of doctors from different specialties will be needed to provide more precise treatment to each patient, and that big data and artificial intelligence will be determinant for the diagnosis, prognosis, treatment, and tracking of illnesses. Better technology will not replace doctors; it will allow them to do a better job.

There is considerable overlap between the terms “precision medicine” and “personalized medicine.” Personalized medicine is an older term whose meaning is similar to precision medicine, but the word “personalized” could be interpreted to mean that treatment and prevention is developed in unique ways for each individual. In precision medicine patients are not treated as unique cases, and medical attention focuses on identifying what approaches are effective for what patients according to their genes, metabolism, lifestyle, and surroundings.

## **Many of today’s tasks will be handled by robots in the future, and teams of doctors from different specialties will provide more precise care to each patient. Big data and artificial intelligence will be determinant for the diagnosis, prognosis, treatment, and tracking of illnesses**

The concept of precision medicine is not new. Patients who receive a transfusion have been matched to donors according to their blood types for over a century, and the same has been occurring with bone marrow and organ transplants for decades. More recently, this has extended to breast-cancer treatment, where the prognosis and treatment decisions are guided primarily by molecular and genetic data about tumorous cells.

Advances in genetics, metabolism, and so on, and the growing availability of clinical data constitute a unique opportunity for making precision-patient treatment a clinical reality. Precision medicine is based on the availability of large-scale data on patients and healthy volunteers. In order for precision medicine to fulfill its promises, hundreds of thousands of people must share their genomic and metabolic data, their health records, and their experiences.

*All of Us* is a research project announced by then President Barack Obama on January 20, 2015, in the following expressive terms: “Tonight, I’m launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes, and to give all of us access to the personalized information we need to keep ourselves and our families healthier.”<sup>2</sup> Three years later, in May 2018, this monumental project started working to recruit over a mil-





lion volunteers to share information about their health (clinical records, biological samples, health surveys, and so on) for many years.

For scientists this represents unique material. The richer our databases are, the more precise our patient care will be. Big data and artificial intelligence are intimately linked; data are the basic components of learning algorithms and with sufficient data and correct analysis they can provide information inconceivable with other techniques. Banks, companies such as Google and Amazon, electric companies, and others have been using big data for decades to improve decision-making and to explore new business opportunities.

Big data's use in science is not new, either. Particle physicists and structural biologists pioneered the development and application of the algorithms behind these technologies. In medicine, big data has the potential to reduce the cost of diagnosis and treatment, to predict epidemics, to help avoid preventable diseases, and to improve the overall quality of life.

For example, various public hospitals in Paris use data from a broad variety of sources to make daily and even hourly predictions of the foreseeable number of patients at each hospital. When a radiologist requests a computerized tomography, artificial intelligence can review the image and immediately identify possible finds based on the image and an analysis of the patient's antecedents with regard to the anatomy being scanned. When a surgeon has to make a complex decision, such as when to operate, whether the intervention will be radical or the organ will be preserved, and to provide precise data as to the potential risks or the probability of greater morbidity or mortality, he or she can obtain such data immediately through artificial intelligence's analysis of a large volume of data on similar interventions.

Of course, various legal and ethical questions have arisen with regard to the conditions under which researchers can access biological samples or data—especially with regard to DNA—as well as the intellectual property derived from their use and questions related to rights of access, confidentiality, publication, and the security and protection of stored data.

The United States' National Institutes of Health (NIH) have shown considerable capacity to resolve difficulties posed by these questions and reach agreements on occasions requiring dialog and effort by all involved parties, including patients, healthy volunteers, doctors, scientists, and specialists in ethical and legal matters.

Genetic, proteomic, and metabolic information, studies of histology and images, microbiomes (the human microbiota consists of the genes stored by these cells), demographic and clinical data, and health surveys are the main registers that constitute the bases for big data.

"Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house."<sup>3</sup> Such was Henri Poincaré's warning that no matter how large a group of data is, it is not therefore science. The Royal Spanish Academy (RAE) defines science as the "body of systematically structured knowledge obtained through observation and reasoning, and from which it is possible to deduce experimentally provable general principles and laws with predictive capacities."<sup>4</sup> The human mind is organized in such a way that it will insist on discovering a relation between any two facts presented to it. Moreover, the greater the distance between them—a symphony by Gustav Mahler and a novel by Thomas Mann, or big data and an illness—the more stimulating the effort to discover their relations.

## Systems Biology

Through the structuring and analysis of databases, scientists seek the basic material for identifying the concepts underlying their confused appearance. One example would be the



effort to establish a precise distinction between two sub-types of the same illness, which share similar histologies but respond differently to treatment or have different prognoses. Systems biology is the branch of biological research that deals with these matters. It seeks to untangle the complexities of biological systems with a holistic approach based on the premise that the interactive networks that make up a living organism are more than the sum of their parts.

Systems biology draws on several disciplines—biochemistry, structural, molecular, and cellular biology, mathematics, biocomputing, molecular imaging, engineering, and so on—to create algorithms for predicting how a biological system changes over time and under different conditions (health vs. illness, fasting vs. eating) and develop solutions for the most pressing health and environmental problems.

On a biological level, the human body consists of multiple networks that combine and communicate with each other at different scales. From our genome to the proteins and metabolites that make up the cells in our organs, we are fundamentally a network of interactive networks that can be defined with algorithms. Systems biology analyzes these networks through their numerous scales to include behavior on different levels, formulating hypotheses about biological functions and providing spatial and temporal knowledge of dynamic biological changes. Studying the complexity of biology requires more than understanding only one part of a system; we have to understand all of it.

The challenge faced by systems biology surpasses any other addressed by science in the past. Take the case of metabolism; the metabolic phenotypes of a cell, organ, or organism are the result of all the catalytic activities of enzymes (the proteins that specifically catalyze a metabolic biochemical reaction) established by kinetic properties, the concentration of substrates, products, cofactors, and so on, and of all the nonlinear regulatory interactions on a transcriptional level (the stage of genetic expression where the DNA sequence is copied in an RNA sequence), a translational level (the process that decodes an RNA sequence to generate a specific chain of amino acids and form a protein), a post-translational level (the covalent modifications that follow a protein's synthesis to generate a great variety of structural and functional changes), and an allosteric level (a change in an enzyme's structure and activity caused by a non-covalent union with another molecule located outside the chemically active center).

In other words, a detailed metabolic description requires knowledge not only of all the factors that influence the quantity and state of enzymes, but also the concentration of all the metabolites that regulate each reaction. Consequently, metabolic flows (the passage of all of the metabolites through all of a system's enzymatic reactions over time) cannot be determined exclusively on the basis of the metabolite concentration, or only by knowing all of the enzymes' nonlinear interactions.

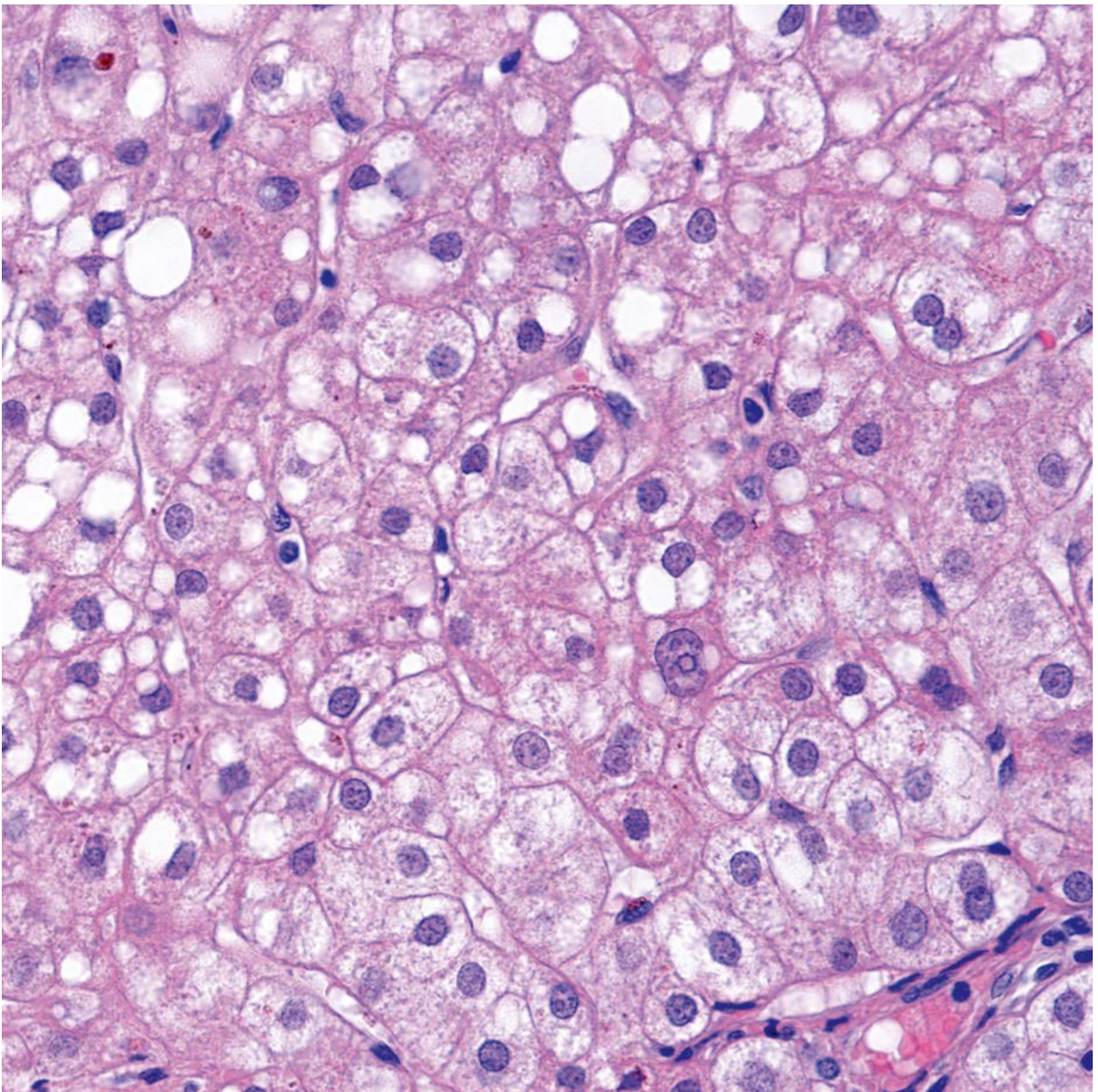
Determining metabolic flows with a certain degree of confidence requires knowing the concentration of all involved metabolites and the quantities of protein in each of the involved enzymes, with their post-translational modifications, as well as a detailed computational model of the reactions they catalyze.

At the same time, a systems-biology approach is needed to integrate large amounts of heterogenous data—transcriptomics (the study of the complete set of RNA transcriptions), proteomics (the study of the complete set of proteins), metabolomics (the study of the complete set of metabolites), and fluxomics (the study in time of the destination of each metabolite and the analysis of the routes they take)—in order to construct an integral metabolic model.

The human genome is estimated to contain between 20,000 and 25,000 genes that codify proteins, of which around 3,000 are enzymes that can be divided into approximately 1,700 metabolic enzymes and 1,300 non-metabolic ones, including those responsible for the post-translational modifications of metabolic enzymes such as kinase proteins, which



Electron micrograph of a liver with macrovascular steatosis (fatty liver). Numerous hepatocytes (liver cells) have one or two large drops of fat that cause swelling and displace the nucleus (dark purple in this image)







**In medicine, big data has the potential to reduce the cost of diagnoses and treatment, to predict epidemics, to avoid preventable diseases, and to improve the overall quality of life**

Detail of breast-cancer cells. This type of cancer affects one in every eight women





In May 2018, former President Obama's *All of Us* program began recruiting over one million volunteers to share information about their health over the course of many years

Supporters of the Patient Protection and Affordable Care Act celebrate its Supreme Court ratification by six votes to three in Washington on June 25, 2015



function as signal transducers to modify the kinetic properties of metabolic enzymes (hence the name kinase, which comes from kinetic) through phosphorylation (bringing phosphoric acid into a protein). These kinase proteins thus play a central role in regulating the metabolism. Metabolic enzymes have been assigned to some 140 different metabolic routes (the set of linked chemical reactions catalyzed by enzymes).

While this human metabolic map could be considered the best-characterized cellular network, it is still incomplete. There are a significant number of metabolites, enzymes, and metabolic routes that have yet to be well characterized or that are simply unknown, even in the oldest metabolic routes, such as glycolysis (the set of chemical reactions that break down certain sugars, including glucose, to obtain energy), the urea cycle (a cycle of five biochemical reactions that turn toxic ammonia into urea to be excreted in urine), and the synthesis of lipids; there are still many gaps in our knowledge.

Moreover, metabolic phenotypes are the product of interactions among a variety of external factors—diet and other lifestyle factors, preservatives, pollutants, environment, microbes—that must also be included in the model.

## **Systems biology draws on several disciplines—biochemistry, structural, molecular and cellular biology, mathematics, biocomputing, molecular imaging, and engineering—to create algorithms for predicting how a biological system changes over time and under different conditions**

The complete collection of small molecules (metabolytes) found in the human body—lipids, amino acids, carbohydrates, nucleic acids, organic acids, biogenic amines, vitamins, minerals, food additives, drugs, cosmetics, pollutants, and any other small chemical substance that humans ingest, metabolize, or come into contact with—is unknown. Around 22,000 have been identified, but the human metabolism is estimated to contain over 80,000 metabolites.

### **Metabolism and Illness**

The idea that illness produces changes in biological fluids, and that chemical patterns can be related to health, is very old. In the middle ages, a series of charts called “urine diagrams” linked the colors, smells, and tastes of urine—all stimuli of a metabolic nature—to diverse medical conditions. They were widely used during that period for diagnosing certain illnesses. If their urine tasted sweet, patients were diagnosed with diabetes, a term that means siphon and refers to the patient’s need to urinate frequently. In 1675, the word *mellitus* (honey) was added to indicate that their urine was sweet; and, in the nineteenth century, methods were developed for detecting the presence of glucose in diabetics’ blood. In 1922, Frederick Banting and his team used a pancreatic extract called “insulin” to successfully treat a patient with diabetes, and he was awarded the Nobel Prize in Medicine the following year. In 1921 Nicolae Paulescu had demonstrated the antidiabetic effect of a pancreatic extract he called “pancreine,” and he patented that discovery in 1922.

The reestablishment of metabolic homeostasis (the set of self-regulatory phenomena that lead to the maintenance of a consistent composition of metabolites and properties inside a



cell, tissue, or organism) through substitution therapy has been successfully employed on numerous occasions. In Parkinson's disease, for example, a reestablishment of dopamine levels through treatment with levodopa has an unquestionable therapeutic effect for which Arvid Carlsson was awarded the Nobel Prize in Medicine in 2000.

An interruption of metabolic flow, however, does not only affect the metabolites directly involved in the interrupted reaction; in a manner that resembles a river's flow, it also produces accumulation of metabolites "upstream." Phenylketonuria, for example, is a rare hereditary metabolic disease (diseases are considered rare when they affect less than one in two thousand people) caused by a phenylalanine hydroxylase deficiency (this enzyme converts phenylalanine into the amino acid called tyrosine) that produces an accumulation of phenylalanine in the blood and brain. At high concentrations, phenylalanine is toxic and causes grave and irreversible anomalies in the brain's structure and functioning. Treatment in newborns affected by this congenital error of the metabolism with a phenylalanine-deficient diet prevents the development of this disease.

Another example is hereditary fructose intolerance (HFI), which appears as a result of a lack of aldolase B, an enzyme that plays a crucial role in the metabolism of fructose and gluconeogenesis—the metabolic route that leads to glucose synthesis from non-carbohydrate precursors. Persons with HFI develop liver and kidney dysfunctions after consuming fructose, which can lead to death, especially during infancy. The only treatment for this rare disease is the elimination of all dietary sources of fructose, including saccharose, fruit juices, asparagus, and peas.

When it is not possible to eliminate a nutrient from the diet to reestablish metabolic homeostasis, it is sometimes possible to design pharmacological chaperones: small molecules that associate with the mutant proteins to stabilize them so they will behave correctly. That is why they are called chaperones.

In the case of certain proteins, such as uroporphyrinogen III synthase, a key enzyme for synthesizing the hemo group (the prosthetic group that is part of numerous proteins, including hemoglobin) whose deficiency causes congenital erythropoietic porphyria, this therapeutic approach has been strikingly successful in an experimental model of the illness, thus supporting the idea that pharmacological chaperones may become powerful therapeutic tools.

## Genetic Therapy and Genome Editing

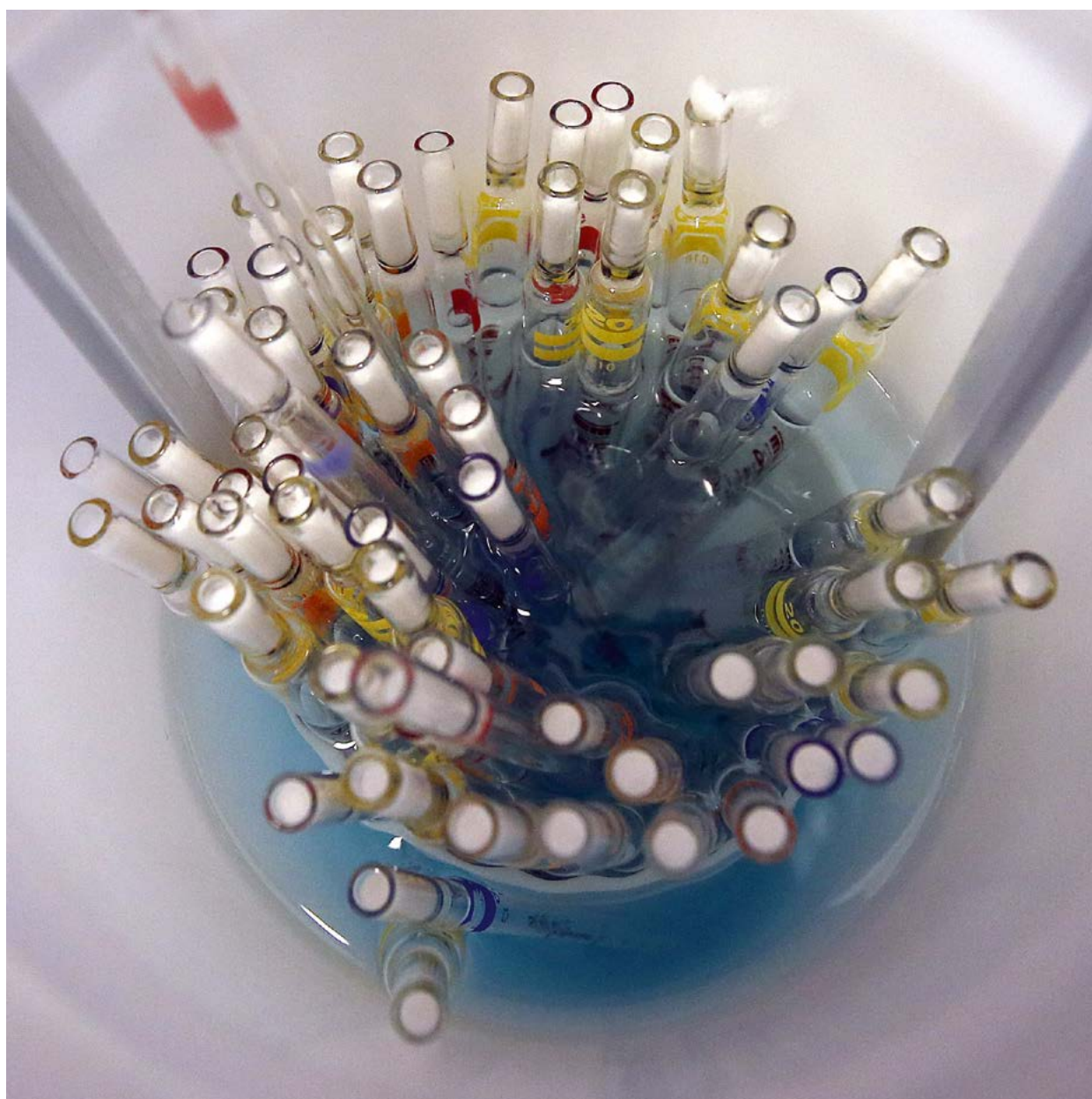
In September 1990, the first genetic therapy (an experimental technique that uses genes to treat or prevent diseases) was carried out at the NIH to combat a four-year-old child's deficiency of the adenosine deaminase enzyme (a rare hereditary disease that causes severe immunodeficiency). In 2016, the European Medicines Agency (EMA) recommended the use of this therapy in children with ADA when they cannot be coupled with a matching bone-marrow donor. While genetic therapy is a promising treatment for various diseases (including hereditary diseases, certain types of cancer, and some viral infections), it continues to be risky thirty years after the first clinical studies, and research continues to make it an effective and safe technique. Currently, it is only being used for illnesses that have no other treatment.

Genome editing is another very interesting technique for preventing or treating illnesses. It is a group of technologies that make it possible to add, eliminate, or alter genetic material at specific locations in the genome. Various approaches to genome editing have been developed. In August 2018, the US Food and Drug Administration agency (FDA) approved the first therapy based on RNA interference (RNAi), a technique discovered twenty years ago that can





Used pipettes in a container await cleaning and disinfection at the Max Planck Institute of Infectious Biology





be used to “silence” the genes responsible for specific diseases. The most recent is known as CRISPR-Cas9, and research is under way to test its application to rare hereditary diseases such as cystic fibrosis or hemophilia, as well as for the treatment and prevention of complex diseases such as cancer, heart disease, or mental illness. Like genetic therapy and RNAi, the medical application of genetic therapy still involves risks that will take time to eliminate.



### Metabolic Transformation

Referring to philology, Friedrich Nietzsche said: “[It] is that venerable art which exacts from its followers one thing above all—to step to one side, to leave themselves spare moments, to grow silent, to become slow—the leisurely art of the goldsmith applied to language: an art which must carry out slow, fine work, and attains nothing if not *lento*.”<sup>5</sup> Replacing *goldsmith* with *smith of the scientific method* will suffice to make this striking observation by Nietzsche applicable to biomedical research. In science, seeking only what is useful quashes the imagination.

In the 1920s, Otto Warburg and his team noticed that, compared to the surrounding tissue, tumors consumed enormous quantities of glucose. They also observed that glucose was metabolized to produce lactate as an energy source (ATP), rather than CO<sub>2</sub>, even when there was enough oxygen for breathing. That process is known as anaerobic glycolysis (none of the nine metabolic routes that join glucose with lactate employs oxygen). This change from aerobic to anaerobic glycolysis in tumors is known as the “Warburg effect” and constitutes a metabolic adaptation that favors the growth of tumor cells. These cells use lactate’s three carbon atoms to synthesize the “basic materials” that make up cells (amino acids, proteins, lipids, nucleic acids, and so on) and generate mass.

We now know that while tumors as a whole exhibit a vast landscape of genetic changes, neoplastic cells exhibit a shared phenotype characterized by disorderly growth that maintains their invasive and lethal potential. Sustaining this unceasing cellular division requires metabolic adaptation that favors the tumor’s survival and destroys normal cells. That adaptation is known as “metabolic transformation.” In fact, although cellular transformation in different types of tumors arises in a variety of ways, the metabolic requirements of the resulting tumor cells are quite similar: cancer cells need to generate energy (ATP), to synthesize “basic materials” to sustain cellular growth and to balance the oxidative stress produced by their ceaseless growth.

## There is currently considerable interest in dissecting the mechanisms and impact of tumors’ metabolic transformation to discover therapies that could block such adaptations and effectively brake cancer growth and metastasis

Due to these shared requirements there is currently considerable interest in dissecting the mechanisms and impact of tumors’ metabolic transformation to discover therapies that could block such adaptations and effectively brake cancer growth and metastasis.

In 1948, Sidney Farber and his team obtained remissions in various children with acute undifferentiated leukemia using aminopterin, an inhibitor of the dihydrofolate reductase enzyme that catalyzes a reaction needed for synthesizing DNA. This molecule was the forerunner of methotrexate, a cancer-treatment drug in common use today.



The age of chemotherapy began with the modification of tumors' metabolisms. Since then, other researchers have discovered molecules capable of slowing the growth and metastasis of various types of cancer, thus reducing annual mortality rates in the United States to an average of 1.4 percent for women and 1.8 percent for men between 1999 and 2015 (Europe shows similar statistics).

In 1976, after analyzing 6,000 fungus samples, Akira Endo isolated mevastatin, a molecule capable of blocking the activity of an enzyme called 3-hydroxy-3-methylglutaryl-coenzyme A reductase, which catalyzes the reaction that limits the synthesis of cholesterol. Mevastatin was the predecessor of lovastatin, a drug in common use since 1984 to treat hypercholesterolemia. Since then, other more efficient and safer statins have been discovered.

The reduction of cholesterol buildups with statins has an unquestionable therapeutic effect on the appearance of cardiovascular incidents. It is one of the principal factors contributing to an over fifty percent drop in the number of deaths from cardiovascular illness and strokes in Western countries over the last few decades. The other two are medication for high blood pressure (a metabolic illness resulting from an imbalance in the homeostasis of water and electrolytes, mainly sodium and potassium) and for diabetes.

## Conclusions

An increase of three decades in the life expectancy of persons born in numerous countries and regions of the West during the twentieth century is the clearest exponent of the social value of science and the triumph of public-health policies over illness: the discovery of vitamins and insulin, the development of vaccines, antibiotics, X-rays, open-heart surgery, chemotherapy, antihypertensive drugs, imaging techniques, antidiabetic drugs, statins, antiviral treatment, immunotherapy, and so on, are examples of the productive collaboration between academia and industry. In 2015, the United Nations took advantage of advances in the Millennium Development Goals to adopt the Sustainable Development Goals, which include the commitment to attain universal health coverage by 2030. However, there are still enormous differences between what can be achieved in health care, and what has actually been attained. Progress has been incomplete and unequal. Meeting this goal requires a full and continuous effort to improve the quality of medical and health-care services worldwide.

Could the process of transforming health knowledge somehow be speeded up? In 1996, James Allison observed that rats treated with an anti-CTLA-4 antibody (a receptor that negatively regulates the response of T lymphocytes) rejected those tumors. Clinical studies began in 2001 and in 2011 the FDA approved this procedure for treating metastatic melanoma, making it the first oncological drug based on activating the immune system. For this work, Allison was awarded the 2018 Nobel Prize for Medicine.

Five years to complete preclinical studies and ten more for clinical trials may seem like a long time; the former answers basic questions about a drug's way of functioning and its safety, although that does not replace studies on how a given drug will interact with the human body. In fact, clinical trials are carried out in stages strictly regulated by drug agencies (EMA, FDA) from initial phase 1 and phase 2 small-scale trials through phase 3 and phase 4 large-scale trials.

Complex design, increased numbers of patients, criteria for inclusion, safety, criteria for demonstrating efficacy, and longer treatment periods are the main reasons behind the longer duration of clinical trials.

It is no exaggeration to state that maintaining the status quo hinders the development of new drugs, especially in the most needy therapeutic areas. Patients, doctors, researchers,

industry, and regulatory agencies must reach agreements that accelerate the progress and precision of treatment. Moreover, the pharmaceutical industry needs to think more carefully about whether the high cost of new treatments puts them out of reach for a significant portion of the general public, with the consequent negative repercussions on quality of life and longevity. Science is a social work.



## Notes

1. WHO, *Health Promotion Glossary*, 1998, p. 16. At <http://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf>

2. This sentence from President Barack Obama's "State of the

Union Address," on January 20, 2015, may be consulted at The White House, Office of the Press Secretary, at <https://obamawhitehouse.archives.gov/the-press-office/2015/01/20/remarks-president-state-union-address-january-20-2015>

3. Henri Poincaré, *Science and Hypothesis*, London and Newcastle-on-Tyne: Walter Scott Publishing, 1905, p. 157.

4. Royal Spanish Academy (RAE), *Diccionario de la Real Academia Española* (DRAE), consulted online at <http://dle.rae.es/?id=9AwuYaT>.

5. Friedrich Nietzsche, *The Dawn of Day*, trans J. McFarland Kennedy, New York: Macmillan, 1911, "Author's Preface," 5 (autumn 1886).



**Daniela Rus**  
MIT-Massachusetts Institute  
of Technology

Daniela Rus is the Andrew (1956) and Erna Viterbi Professor of Electrical Engineering and Computer Science and Director of the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT. Rus's research interests are in robotics, artificial intelligence, and data science. The focus of her work is developing the science and engineering of autonomy, toward the long-term objective of enabling a future with machines pervasively integrated into the fabric of life, supporting people with cognitive and physical tasks. Her research addresses some of the gaps between where robots are today and the promise of pervasive robots: increasing the ability of machines to reason, learn, and adapt to complex tasks in human-centered environments, developing intuitive interfaces between robots and people, and creating the tools for designing and fabricating new robots quickly and efficiently. The applications of this work are broad and include transportation, manufacturing, agriculture, construction, monitoring the environment, underwater exploration, smart cities, medicine, and in-home tasks such as cooking. Rus serves as the Associate Director of MIT's Quest for Intelligence Core, and as Director of the Toyota-CSAIL Joint Research Center, whose focus is the advancement of AI research and its applications to intelligent vehicles. She is a member of the Toyota Research Institute advisory board. Rus is a Class of 2002 MacArthur Fellow, a fellow of ACM, AAAI, and IEEE, and a member of the National Academy of Engineering and the American Academy of Arts and Sciences. She is the recipient of the 2017 Engelberger Robotics Award from the Robotics Industries Association. She earned her PhD in Computer Science from Cornell University.

Recommended book: *Machines of Loving Grace*, John Markoff, Harper Collins Publishers, 2015.

**By customizing and democratizing the use of machines, we bring robots into the forefront. Pervasive integration of robots in the fabric of everyday life may mean that everyone could rely on a robot to support their physical tasks, just like we have come to rely on applications for computational tasks. As robots move from our imaginations into our homes, offices, and factory floors, they will become the partners that help us do so much more than we can do alone. Whether in how we move, what we build, where we do it, or even the fundamental building blocks of creation, robotics will enable a world of endless possibility.**





Imagine a future where robots are so integrated in the fabric of human life that they become as common as smartphones are today. The field of robotics has the potential to greatly improve the quality of our lives at work, at home, and at play by providing people with support for cognitive tasks and physical tasks. For years, robots have supported human activity in dangerous, dirty, and dull tasks, and have enabled the exploration of unreachable environments, from the deep oceans to deep space. Increasingly, more capable robots will be able to adapt, to learn, and to interact with humans and other machines at cognitive levels.

The rapid progress of technology over the past decade has made computing indispensable. Computing has transformed the way we work, live, and play. The digitization of practically everything, coupled with advances in robotics, promises a future where access to high-tech machines is democratized and customization widespread. Robots are becoming increasingly capable due to their ability to execute more complex computations and interact with the world through increasingly richer sensors and better actuators.

A connected world with many customized robots working alongside people is already creating new jobs, improving the quality of existing jobs, and saving people time so they can focus on what they find interesting, important, and exciting. Today, robots have already become our partners in industrial and domestic settings. They work side by side with people in factories and operating rooms. They mow our lawns, vacuum our floors, and even milk our cows. In a few years, they will touch even more parts of our lives.

Commuting to work in your driverless car will let you read, return calls, catch up on your favorite podcasts, and even nap. The robotic car will also serve as your assistant, keeping track of what you need to do, planning your routes to ensure all your chores are done, and checking the latest traffic data to select the least congested roads. Driverless cars will help reduce fatalities from car accidents while autonomous forklifts can help eliminate back injuries caused by lifting heavy objects. Robots may change some existing jobs, but, overall, robots can make great societal contributions. Lawn-care robots and pool-cleaning robots have changed how these tasks are done. Robots can assist humanity with problems big and small.

## **The digitization of practically everything, coupled with advances in robotics, promises a future where access to high-tech machines is democratized and customization widespread**

The objective of robotics is not to replace humans by mechanizing and automating tasks, but, rather, to find new ways that allow robots to collaborate with humans more effectively. Robots are better than humans at tasks such as crunching numbers and moving with precision. Robots can lift much heavier objects. Humans are better than robots at tasks like reasoning, defining abstractions, and generalizing or specializing thanks to our ability to draw on prior experiences. By working together, robots and humans can augment and complement each other's skills.

### **A Decade of Progress Enabling Autonomy**

The advancements in robotics over the past decade have demonstrated that robotic devices can locomote, manipulate, and interact with people and their environment in unique ways.



The locomotion capabilities of robots have been enabled by the wide availability of accurate sensors (for example, laser scanners), high-performance motors, and development of robust algorithms for mapping, localization, motion planning, and waypoint navigation. Many new applications are possible thanks to progress in developing robot bodies (hardware) and robot brains (software).

The capabilities of robots are defined by the tight coupling between their physical bodies and the computation that comprises their brains. For example, a flying robot must have a body capable of flying and algorithms to control flight. Today's robots can do basic locomotion on the ground, in the air, and in water. They can recognize objects, map new environments, perform pick-and-place operations, learn to improve control, imitate simple human motions, acquire new knowledge, and even act as a coordinated team. For example, the latest soccer robots and algorithms are put in practice at a yearly robot soccer competition called RoboCup.

Recent advances in disk storage, the scale and performance of the Internet, wireless communication, tools supporting design and manufacturing, and the power and efficiency of electronics, coupled with the worldwide growth of data storage, have impacted the development of robots in multiple ways. Hardware costs are going down, electromechanical components are more reliable, tools for making robots are richer, programming environments are more readily available, and robots have access to the world's knowledge through the cloud. We can begin to imagine the leap from the personal computer to the personal robot, leading to many applications where robots exist pervasively and work side by side with humans.

## **The objective of robotics is not to replace humans by mechanizing and automating tasks, but, rather, to find new ways that allow robots to collaborate with humans more effectively**

Transportation is a great example. It is much easier to move a robot through the world than it is to build a robot that can interact with it. Over the last decade, significant advances in algorithms and hardware have made it possible for us to envision a world in which people and goods are moved in a much safer, more convenient way with optimized fleets of self-driving cars.

In a single year, Americans drive nearly three trillion miles.<sup>1</sup> If you average that out at 60 mph, that adds up to almost fifty billion hours spent in the car.<sup>2</sup> That number grows exponentially when considering the rest of the globe. But the time spent in our cars is not without challenges. Today, a car crash occurs every five seconds in the United States.<sup>3</sup> Globally, road traffic injuries are the eighth leading cause of death, with about 1.24 million lives lost every year.<sup>4</sup> In addition to this terrible human cost, these crashes take an enormous economic toll. The National Highway Traffic Safety Administration has calculated the economic cost in the United States at about \$277 billion a year.<sup>5</sup> Putting a dent in these numbers is an enormous challenge—but one that is very important to tackle. Self-driving vehicles have the potential to eliminate road accidents.

Imagine if cars could learn... learn how we drive... learn how to never be responsible for a collision... learn what we need when we drive? What if they could become trusted partners? Partners that could help us navigate tricky roads, watch our backs when we are tired, even make our time in the car... fun? What if your car could tell you are having a hard day and



turn your favorite music on to help you relax, while watching carefully over how you drive? What if your car also knew that you forgot to call your parents yesterday and issued a gentle reminder on the way home. And imagine that it was easy to make that call because you could turn the driving over to the car on a boring stretch of highway.

Recognizing this extraordinary potential during the past couple of years, most car manufacturers announced self-driving car projects. Elon Musk famously predicted we could fall asleep at the wheel in five years; the Google/Waymo car has been in the news a lot for driving several million accident-free miles; Nissan promised self-driving cars by 2020; Mercedes created a prototype 2014 Model S Autonomous car; and Toyota announced (September 2015) an ambitious program to develop a car that will never be responsible for a collision, and invested \$1 billion to advance artificial intelligence.

There is a lot of activity in this space across a big spectrum of car capabilities. To understand where all the various advances fall, it is useful to look at the National Highway Traffic Safety Administration (NHTSA) classification of five levels of autonomy: Level 0 does not include any support for automation; Level 1 includes tools for additional feedback to the human driver, for example using a rear camera; Level 2 includes localized active control, for example antilock brakes; Level 3 includes support for select autonomy but the human must be ready to take over (as in the Tesla Autopilot); Level 4 includes autonomy in some places some of the time; and Level 5 is autonomy in all environments all the time.

An alternative way to characterize the level of autonomy of a self-driving car is according to three axes defining (1) the speed of the vehicle; (2) the complexity of the environment in which the vehicle moves, and (3) the complexity of the interactions with moving agents (cars, people, bicyclists, and so on) in that environment. Researchers are pushing the envelope along each of these axes, with the objective to get closer to Level 5 autonomy.

## **Over the last decade, significant advances in algorithms and hardware have made it possible for us to envision a world in which people and goods are moved in a much safer, more convenient way with optimized fleets of self-driving cars**

Due to algorithmic and hardware advances over the past decade, today's technology is ready for Level 4 deployments at low speeds in low-complexity environments with low levels of interaction with surrounding pedestrians and other vehicles. This includes autonomy on private roads, such as in retirement communities and campuses, or on public roads that are not very congested.

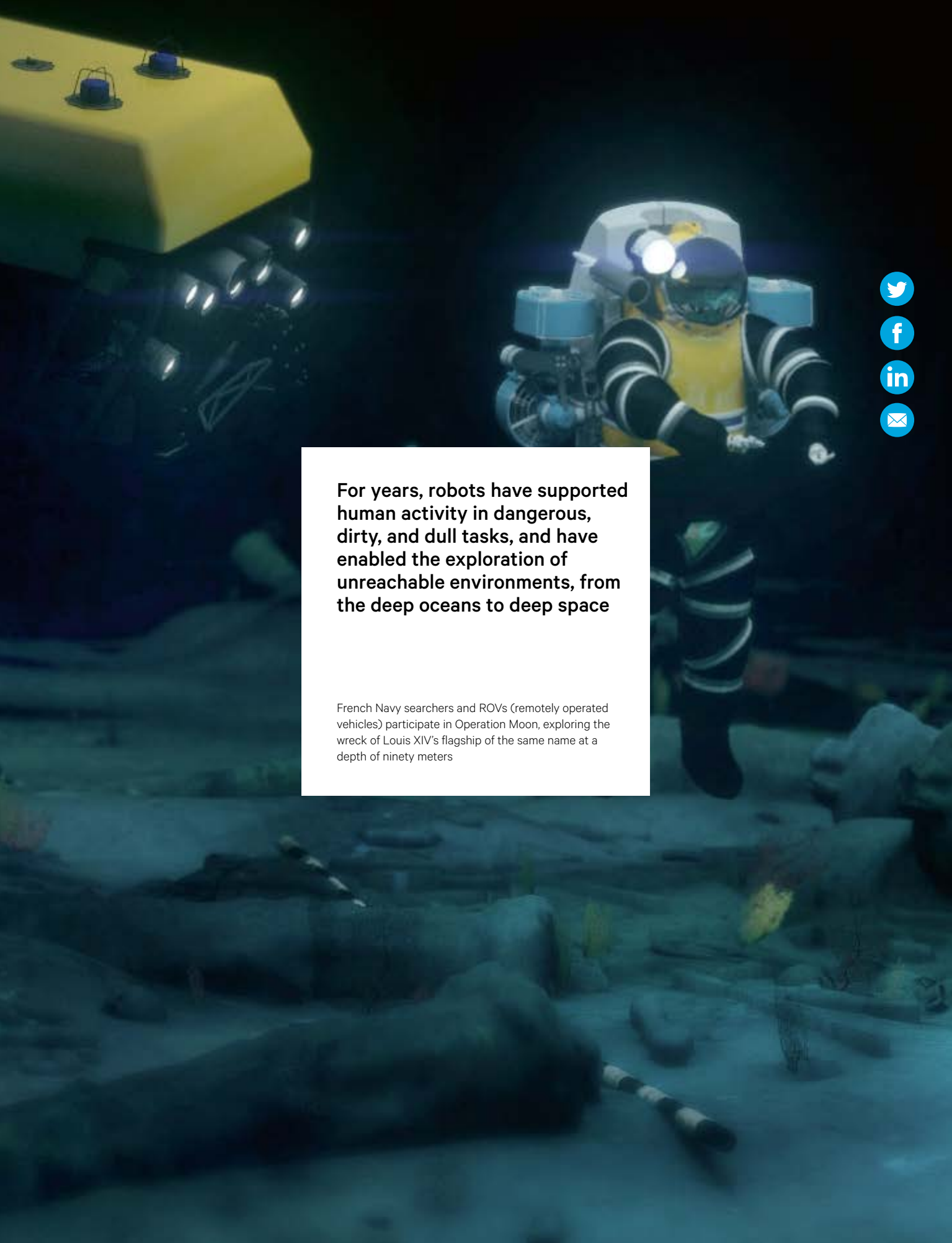
Level 4 autonomy has been enabled by a decade of advances in the hardware and algorithms available to the robots. Most important is the convergence of several important algorithmic developments: map making, meaning the vehicle can use its sensors to create a map; localization, meaning the vehicle can use its sensors to figure out where it is on the map; perception, meaning the vehicle can perceive the moving objects on the road; planning and decision-making, meaning the vehicle can figure out what to do next based on what it sees now; and reliable hardware, as well as driving datasets that enable cars to learn how to drive from humans. Today, we can do so many simultaneous computations, crunch so much more data, and run algorithms in real time. These technologies have taken us to a point in time where we can realistically discuss the idea of autonomy on the roads.



The Da Vinci surgical robot during a hysterectomy operation



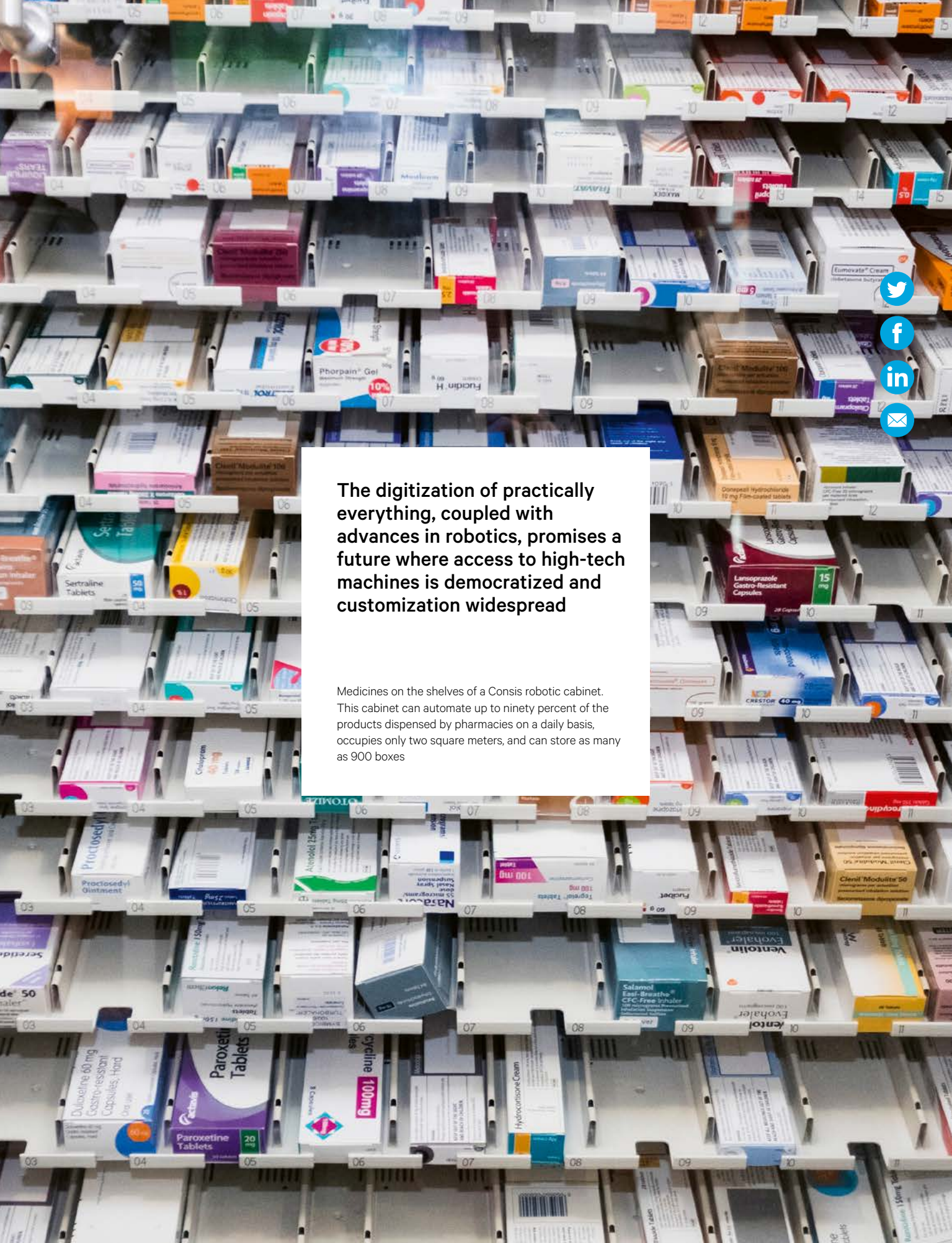




**For years, robots have supported human activity in dangerous, dirty, and dull tasks, and have enabled the exploration of unreachable environments, from the deep oceans to deep space**

French Navy searchers and ROVs (remotely operated vehicles) participate in Operation Moon, exploring the wreck of Louis XIV's flagship of the same name at a depth of ninety meters



A robotic pharmacy cabinet with numerous shelves, each holding a box of medicine. The shelves are labeled with numbers from 01 to 10. The boxes are of various colors and sizes, representing different medications. In the center of the image, there is a white text box with black text. On the right side of the image, there are four circular social media icons: Twitter, Facebook, LinkedIn, and Email.

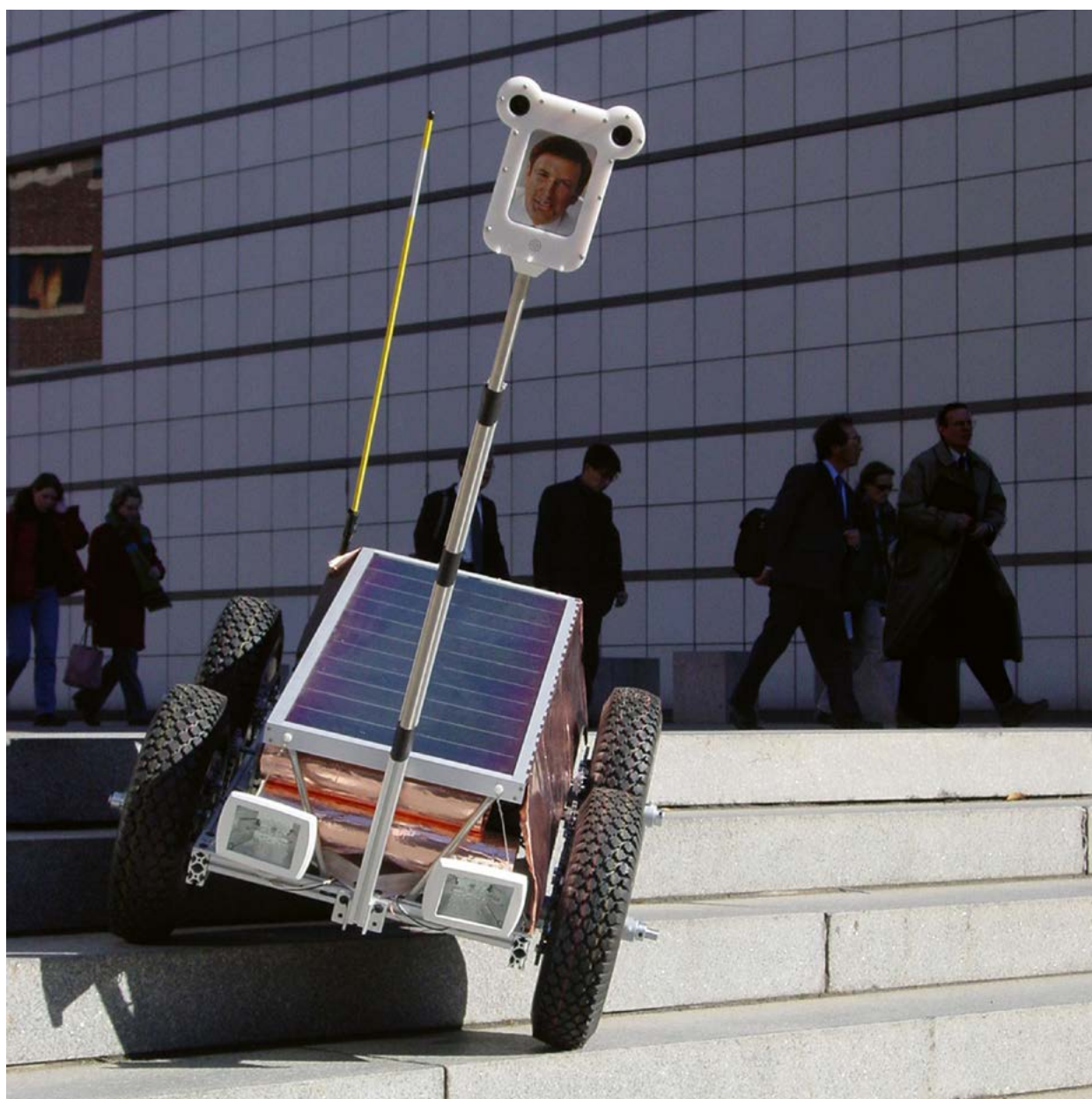
**The digitization of practically everything, coupled with advances in robotics, promises a future where access to high-tech machines is democratized and customization widespread**

Medicines on the shelves of a Consis robotic cabinet. This cabinet can automate up to ninety percent of the products dispensed by pharmacies on a daily basis, occupies only two square meters, and can store as many as 900 boxes





Afghan eXplorer, a semiautomatic mobile robot developed by the Artificial Intelligence Lab of the Massachusetts Institute of Technology (MIT), can carry out reporting activities in dangerous or inaccessible surroundings





However, we do not have Level 5 autonomy yet. Technological challenges toward Level 5 autonomy include: driving in congestion, driving at high speeds, driving in inclement weather (rain, snow), driving among human drivers, driving in areas where there are no high-density maps, and responding to corner cases. The perception system of a vehicle does not have the same quality and effectiveness as the human eye. To be clear, there are some things that machines can do better than people, like estimate accurately how quickly another vehicle is moving. But robots do not share our recognition capabilities. How could they? We spend our whole lives learning how to observe the world and make sense of it. Machines require algorithms to do this, and data—lots and lots and lots of data, annotated to tell them what it all means. To make autonomy possible, we have to develop new algorithms that help them learn from far fewer examples in an unsupervised way, without constant human intervention.

There are two philosophies that are driving research and development in autonomous driving: series autonomy and parallel autonomy. Parallel autonomy concerns developing driver-assist technologies where the driver is at the wheel, but the driver-assist system monitors what the driver does and intervenes as needed—in a way that does no harm—for example to prevent a collision or to correct the steering angle that keeps the car on the road. The autonomy capabilities of the car grow incrementally but operate in parallel with the human. The parallel autonomy approach allows the car to operate anywhere, anytime. Series autonomy pursues the idea that the human or the car are in charge, but not both. When the car is in autonomous mode, the human does not contribute in any way to the driving. The car's autonomy capabilities also grow incrementally, but this car can only operate according to the capabilities supported by its autonomy package. The car will gradually operate in increasingly more complex environments.

**There are two philosophies driving research and development in autonomous driving: series autonomy and parallel autonomy. The latter concerns developing driver-assist technologies where the driver is at the wheel, but the driver-assist system monitors what the driver does and intervenes as needed**

Today's series autonomy solutions operate in closed environments (defined by the roads on which the vehicle can drive). The autonomy recipe starts by augmenting the vehicles with drive-by-wire control and sensors such as cameras and laser scanners. The sensors are used to create maps, to detect moving obstacles, such as pedestrians and other vehicles, and to localize the vehicle in the world. The autonomous driving solutions are map-based and benefit from a decade of progress in the area of simultaneous localization and mapping (SLAM). The maps are constructed by driving the autonomous vehicle on every possible road segment, collecting features with the sensors. The maps are used for each subsequent autonomous drive, to plan a path from start to goal, to execute the path while avoiding obstacles, and to localize the vehicles as it executes the path.

Most self-driving car companies only test their fleets in major cities where they have developed detailed 3D maps that are meticulously labeled with the exact positions of things like lanes, curbs, and stop signs. These maps include environmental features detected by the sensors of the vehicle. The maps are created using 3D LIDAR systems that rely on light





to scan the local space, accumulating millions of data points and extracting the features defining each place.

If we want self-driving cars to be viable global technology, this reliance on detailed prior maps is a problem. Today's autonomous vehicles are not able to drive in rural environments where we do not have maps—in other words, on the millions of miles of roads that are unpaved, unlit, or unreliably marked. At the MIT CSAIL, we began developing MapLite as a first step for enabling self-driving cars to navigate on roads that they have never been on before using only GPS and sensors. Our system combines GPS data—like the kind you would find on Google Maps—with data taken from LIDAR sensors. Together, these two elements allow us to autonomously drive a car on multiple unpaved country roads and reliably detect the road more than 100 feet (30 meters) in advance. Other researchers have been working on different map-less approaches with varying degrees of success. Methods that use perception sensors like LIDAR often have to rely heavily on road markings or make broad generalizations about the geometry of road curbs. Meanwhile, vision-based approaches can perform well in ideal conditions, but have issues when there is adverse weather or bad lighting. In terms of “Level 5 autonomy”—that is, autonomy anywhere any time—we are still some years away, and this is because of both technical and regulatory challenges.

## **Autonomous vehicles can take many different forms, including golf carts, wheelchairs, scooters, luggage, shopping carts, garbage bins and even boats. These technologies open the door to a vast array of new products and applications**

While progress has been significant on the technical side, getting policy to catch up has been an understandably complex and incremental process. Policy makers are still debating the level at which autonomous vehicles should be regulated. What kinds of vehicles should be allowed on the road, and who is allowed to operate them? How should safety be tested, and by whom? How might different liability regimes shape the timely and safe adoption of autonomous vehicles, and what are the trade-offs? What are the implications of a patchwork of state-by-state laws and regulations, and what are the trade-offs in harmonizing these policies? To what extent should policy makers encourage the adoption of autonomous vehicles? For example, through smart-road infrastructure, dedicated highway lanes, manufacturer or consumer incentives? These are complex issues regarding the use of autonomous vehicles on public roads. At the same time, a form of autonomy that is already deployable now is “Level 4 autonomy,” defined as autonomy in some environments some of the time. The technology is here for autonomous vehicles that can drive in fair weather, on private ways, and at lower speeds.

Environments such as retirement communities, campuses, hotel properties, and amusement parks can all benefit from the Level 4 autonomy technologies. Autonomous vehicles can take many different forms, including golf carts, wheelchairs, scooters, luggage, shopping carts, garbage bins, and even boats. These technologies open the door to a vast array of new products and applications, from mobility on demand, to autonomous shopping and transportation of goods, and more efficient mobility in hospitals. Everyone would benefit from transportation becoming a widely available utility, but those benefits will have a particular impact on new drivers, our senior population, and people affected by illness or disability.

The technology that is enabling autonomy for cars can have a very broad societal impact. Imagine residents of a retirement community being transported safely by automated golf carts. In the future, we will be able to automate anything on wheels—not just the vacuum cleaners of today, but also lawn mowers or even garbage cans.

## **If we want self-driving cars to be viable global technology, this reliance on detailed prior maps is a problem. Today's autonomous vehicles are not able to drive in rural environments where we do not have maps**

The same technology that will enable this level of automation could even be put to use to help people dealing with disabilities—like the blind—experience the world in ways never before possible. Visual impairment affects approximately 285 million people worldwide, people who could benefit enormously from increased mobility and robotic assistance. This is a segment of the population that technology has often left behind or ignored, but, in this case, technology could make all the difference. Wearable devices that include the sensors used by self-driving cars and run autonomy software could enable visually impaired people to experience the world safely and in ways that are much richer than the walking stick.

Robotics will change the way we transport people and things in the very near future. But, soon after, it will do more than deliver things on time; it will also enable us to produce those things quickly and locally.

### **Challenges in Robotics**

Despite recent and significant strides in the field, and promise for the future, today's robots are still quite limited in their ability to figure things out, their communication is often brittle, and it takes too much time to make new robots. Broad adoption of robots will require a natural integration of robots in the human world rather than an integration of humans into the machines' world.

**Reasoning** Robots can only perform limited reasoning due to the fact that their computations are carefully specified. For today's robots, everything is spelled out with simple instructions and the scope of the robot is entirely contained in its program. Tasks that humans take for granted, for example asking the question “Have I been here before?” are notoriously difficult for robots. Robots record the features of the places they have visited. These features are extracted from sensors such as cameras or laser scanners. It is hard for a machine to differentiate between features that belong to a scene the robot has already seen and a new scene that happens to contain some of the same objects. In general, the data collected from sensors and actuators is too big and too low level; it needs to be mapped to meaningful abstractions for robots to be able to effectively use the information. Current machine-learning research on big data is addressing how to compress a large dataset to a small number of semantically meaningful data points. Summarization can also be used by robots. For example, robots could summarize their visual history to reduce significantly the number of images required to determine whether “I have been here before.”

Additionally, robots cannot cope with unexpected situations. If a robot encounters a case it was not programmed to handle or is outside the scope of its capabilities, it will enter an error state and halt. Often the robot cannot communicate the cause of the error. For example,





vacuum-cleaning robots are designed and programmed to move on the floor, but cannot climb stairs.

Robots need to learn how to adjust their programs, adapting to their surroundings and the interactions they have with people, with their environments and with other machines. Today, everybody with Internet access has the world's information at their fingertips, including machines. Robots could take advantage of this information to make better decisions. Robots could also record and use their entire history (for example, output of their sensors and actuators), and the experiences of other machines. For example, a robot trained to walk your dog could access the weather report online, and then, based on previous walks, determine the best route to take. Perhaps a short walk if it is hot or raining, or a long walk to a nearby park where other robotic dog walkers are currently located. All of this could be determined without human interaction or intervention.

**Communication** A world with many robots working together requires reliable communication for coordination. Despite advances in wireless communication, there are still impediments in robot-to-robot communication. The problem is that modeling and predicting communication is notoriously hard and any robot control method that relies on current communication models is fraught with noise. The robots need more reliable approaches to communication that guarantee the bandwidth they need, when they need it. To get resilient robot-to-robot communication, a new paradigm is to measure locally the communication quality instead of predicting it with models. Using the idea of measuring communication, we can begin to imagine using flying robots as mobile base-stations that coordinate with each other to provide planet-scale communication coverage. Swarms of flying robots could bring Internet access everywhere in the world.

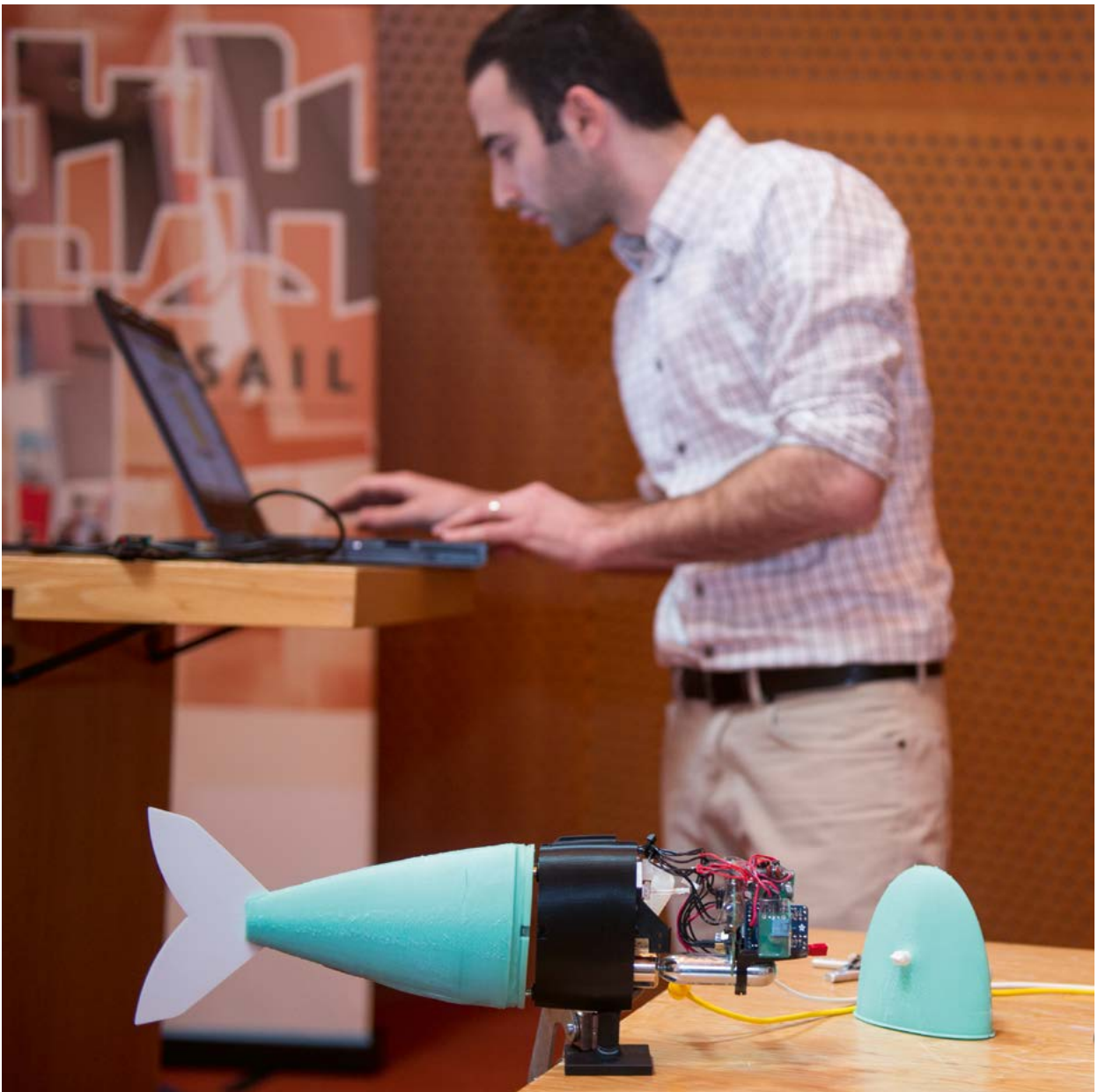
Communication between robots and people is also currently limited. While speech technologies have been employed to give robots commands in human language (for example, "move to the door"), the scope and vocabulary of these interactions is shallow. Robots could use the help of humans when they get stuck. It turns out that even a tiny amount of human intervention in the task of a robot completely changes the problem and empowers the machines to do more.

Currently, when robots encounter something unexpected (a case for which it was not programmed) they get stuck. Suppose, instead of just getting stuck, the robot was able to reason about why it is stuck and enlist human help. For example, recent work on using robots to assemble IKEA furniture demonstrates that robots can recognize when a table leg is out of reach and ask humans to hand them the part. After receiving the part, the robots resume the assembly task. These are some of the first steps toward creating symbiotic human-robot teams where robots and humans can ask each other for help.

**Design and Fabrication** Another great challenge with today's robots is the length of time to design and fabricate new robots. We need to speed up the creation of robots. Many different types of robots are available today, but each of these robots took many years to produce. The computation, mobility, and manipulation capabilities of robots are tightly coupled to the body of the robot—its hardware system. Since today's robot bodies are fixed and difficult to extend, the capabilities of each robot are limited by its body. Fabricating new robots—add-on robotic modules, fixtures, or specialized tools to extend capabilities—is not a real option, as the process of design, fabrication, assembly, and programming is long and cumbersome. We need tools that will speed up the design and fabrication of robots. Imagine creating a robot compiler that takes as input the functional specification of the robot (for example



University student Andrew Marchese demonstrates the movement of a robotic fish during a show at MIT's Artificial Intelligence Lab in April 2013. The robotic fish simulates the movement of living fish and employs the emerging field of soft robotics





“I want a robot to play chess with me”) and computes a design that meets the specification, a fabrication plan, and a custom-programming environment for using the robot. Many tasks big and small could be automated by rapid design and fabrication of many different types of robots using such a robot compiler.

**Toward Pervasive Robotics** There are significant gaps between where robots are today and the promise of pervasive integration of robots in everyday life. Some of the gaps concern the creation of robots—how do we design and fabricate new robots quickly and efficiently? Other gaps concern the computation and capabilities of robots to reason, change, and adapt for increasingly more complex tasks in increasingly complex environments. Other gaps pertain to interactions between robots, and between robots and people. Current research directions in robotics push the envelope in each of these directions, aiming for better solutions to making robots, controlling the movement of robots and their manipulation skills, increasing the ability for robots to reason, enabling semantic-level perception through machine vision, and developing more flexible coordination and cooperation between machines and between machines and humans. Meeting these challenges will bring robots closer to the vision of pervasive robotics: the connected world of many people and many robots performing many different tasks.

**There are significant gaps between where robots are today and the promise of pervasive integration of robots in everyday life. These gaps concern the creation of robots, their computation and capacity to reason, change, and adapt for increasingly more complex tasks in increasingly complex environments, and their capacity to interact with people**

Pervasive, customized robotics is a big challenge, but its scope is not unlike the challenge of pervasive computing, which was formulated about twenty-five years ago. Today we can say that computing is indeed pervasive, it has become a utility, and is available anywhere, anytime. So, what would it take to have pervasive integration of robots in everyday life? Mark Weiser, who was a chief scientist at Xerox PARC and is widely referred to as the father of ubiquitous computing, said of pervasive computing that: “The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.”

For example, electricity was once a novel technology and now it is a part of life. Robotic technologies have the potential to join the personal computer and electricity as pervasive aspects of everyday life. In the near future, robotic technologies will change how we think about many aspects of everyday life.

Self-driving car fleets have the potential to turn transportation into a utility, with customized rides available anywhere, anytime. Public transportation could become a two-layer system: a network of large vehicles (for example, trains, buses) providing backbone transportation for many people over long distances, and fleets of transportation pods providing the customized transportation needs of individuals for short hops. Such a transportation network would be connected to the IT infrastructure and to people to provide mobility on demand. The



operation of the backbone could include dynamically changing routes to adapt to people's needs. Real-time and historical transportation data are already used to determine the most optimal bus routes and location of stops at a fine granularity. Mobility on demand may be facilitated by state-of-the-art technologies for self-driving vehicles. Taking a driverless car for a ride could be as easy as using a smartphone. The robot pods would know when people arrive at a station, where the people are who need a ride now, and where the other robot pods are. After driving people to their destination, the robot pods would drive themselves to the next customer, using demand-matching and coordination algorithms to optimize the operations of the fleet and minimize people's waiting time. Public transportation would be convenient and customized.



## Notes

1. [http://www.fhwa.dot.gov/policyinformation/travel\\_monitoring/14augvt/page2.cfm](http://www.fhwa.dot.gov/policyinformation/travel_monitoring/14augvt/page2.cfm).
2. Exact figure is 496 billion.  
<http://www.forbes.com/sites/modeledbehavior/2014/11/08/the-massive-economic-benefits-of-self-driving-cars/>.
3. [https://www.osha.gov/Publications/motor\\_vehicle\\_guide.pdf](https://www.osha.gov/Publications/motor_vehicle_guide.pdf).
4. [http://apps.who.int/iris/bitstream/10665/83789/1/WHO\\_NMH\\_VIP\\_13.01\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/83789/1/WHO_NMH_VIP_13.01_eng.pdf?ua=1).
5. From autonomous driving ppt.



**Samuel H. Sternberg**  
Columbia University

Samuel H. Sternberg, PhD, runs a research laboratory at Columbia University, where he is an Assistant Professor in the Department of Biochemistry and Molecular Biophysics. He received his BA in Biochemistry from Columbia University in 2007, graduating *summa cum laude*, and his PhD in Chemistry from the University of California, Berkeley, in 2014. He earned graduate student fellowships from the National Science Foundation and the Department of Defense, and received the Scaringe Award and the Harold Weintraub Graduate Student Award. Sam's research focuses on the mechanism of DNA targeting by RNA-guided bacterial immune systems (CRISPR-Cas) and on the development of these systems for genome engineering. In addition to publishing his work in leading scientific journals, he recently coauthored a popular science trade book together with Jennifer Doudna, entitled *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*, about the discovery, development, and applications of CRISPR gene-editing technology.

Recommended book: *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*, J. A. Doudna and S. H. Sternberg, Houghton Mifflin Harcourt, 2017.

**Few discoveries transform a discipline overnight, but scientists today can manipulate cells in ways hardly imaginable before, thanks to a peculiar technology known as CRISPR (clustered regularly interspaced short palindromic repeats). From elegant studies that deciphered how CRISPRs function in bacteria, researchers quickly uncovered the biological potential of Cas9, an RNA-guided DNA cleaving enzyme, for gene editing. Today, this core capability is being harnessed for a wide variety of ambitious applications, including agricultural improvement, the elimination of infectious diseases, and human therapeutics. CRISPR technology may indeed herald cures to certain genetic diseases and cancer, but so too could it be used to engineer heritable genetic changes in human embryos. What will we choose to do with this awesome power?**



Ever since the discovery of DNA as the genetic information carrier of the cell, scientists have been on an inexorable quest to decrypt the letter-by-letter sequence of the human genome, and to develop tools to manipulate and modify genetic code. The first viral genome was fully sequenced in 1976, and, in the years since, an increasing number of ever-more challenging genomes have been decoded, culminating with the first draft of the human genome published in 2001. Today, interested individuals can purchase genetic tests and learn about their ancestry and disease susceptibility, simply by sending saliva samples in the mail, and a growing number of companies are leveraging huge datasets from millions of individuals to rapidly advance the field of human genomics.

Concurrently, tools to build synthetic DNA molecules in the lab have expanded considerably, beginning with the recombinant DNA revolution in the 1970s. In 2010, researchers at the Venter Institute succeeded in creating the first synthetic cell by manufacturing an entire bacterial genome from scratch, and others have since moved on to build entire chromosomes—the cellular structures that contain DNA—in yeast.

Yet, up until quite recently, tools to *modify* DNA, and to do so *directly in living cells*, lagged far behind, particularly in more complex organisms like plants, animals, and humans. In light of our growing awareness of the large number of diseases that are associated with, or caused by, genetic mutations—sickle cell, Huntington's, and Alzheimer's, to name just a few—the ideal DNA manipulation tool would enable direct repair of mutations at their source. But the sheer size and complexity of our genome, made up of 3.2 billion letters of DNA contributed by each of the twenty-three pairs of chromosomes we inherit from mom and dad, consigned the dream of precision genome editing to a distant future. Until, that is, CRISPR technology came along.

Today, scientists can use CRISPR to engineer the genome in ways barely imaginable before: repairing genetic mutations, removing pathogenic DNA sequences, inserting therapeutic genes, turning genes on or off, and more. CRISPR democratized genome engineering, unlike the technologies that preceded it, because it is easy to deploy and inexpensive to access. And CRISPR works in an impressive number of different cell types and organisms—everything from maize to mice to monkeys—giving scientists a flexible, diverse, and broadly applicable engineering toolkit to address a wide variety of biological challenges.

## Scientists can use CRISPR to engineer the genome in ways barely imaginable before: repairing genetic mutations, removing pathogenic DNA sequences, inserting therapeutic genes, turning genes on or off, and more

What is CRISPR, and how does it work? What is meant by gene editing? Are CRISPR-based treatments to cure genetic disease around the corner? How can CRISPR technology be used to improve agriculture through plant and animal gene editing? Could CRISPR help eradicate pathogenic diseases like malaria? And, perhaps most profoundly, might CRISPR someday be used to rewrite DNA in human embryos, thereby editing genetic code in a way that would be felt for generations to come?





As revolutionary as CRISPR has been for biomedical science, its discovery stemmed from basic scientific curiosity about a biological topic about as far removed from medicine as it gets. To understand where CRISPR comes from, we need to delve into one of the longest standing genetic conflicts on Earth: the relentless arms race between bacteria and bacteria-specific viruses (Rohwer et al., 2014).

Everyone knows about bacteria, those pesky microorganisms that can make us sick—think *Streptococci*, the cause of strep throat and pneumonia, or *Salmonella* infections that cause food poisoning—but which are also indispensable for normal human function. (We depend on a vast army of bacteria that collectively make up our microbiome and help break down food, produce vitamins, and perform numerous other essential functions.) Few outside the research community, though, may know about the ubiquity of bacterial viruses, also known as bacteriophages (“eaters of bacteria”). In fact, bacteriophages are by far the most prevalent form of life on our planet: at an estimated abundance of ten million trillion trillion, they outnumber even bacteria ten to one. There are approximately one trillion bacterial viruses for every grain of sand in the world, and ten million viruses in every drop of seawater (Keen, 2015)!

Bacterial viruses evolved to infect bacteria, and they do so remarkably well. They exhibit three-dimensional structures that are exquisitely well suited to latch onto the outer surface of bacterial cells, and after attaching themselves in this manner, they inject their genetic material inside the bacterial host using pressures similar to that of an uncorked champagne bottle. After the viral genome makes its way inside the bacteria, it hijacks the host machinery to replicate its genetic code and build more viruses, ultimately destroying the cell in the process. Roughly twenty to forty percent of the ocean’s bacteria are eliminated every day from such viral infections, vastly reshaping the marine ecosystem by causing release of carbon and other nutrients back into the environment.

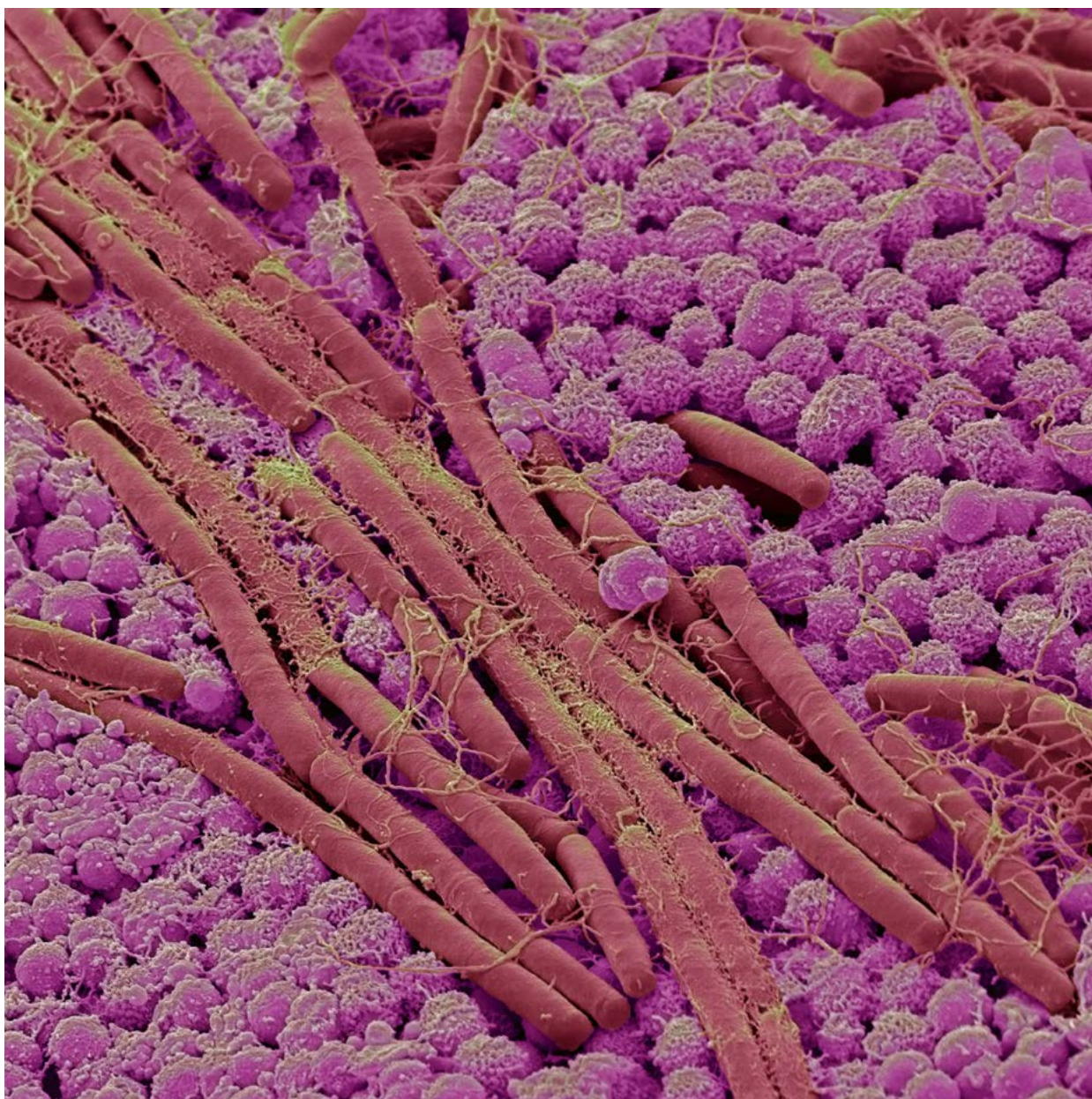
## To understand where CRISPR comes from, we need to delve into one of the longest standing genetic conflicts on Earth: the relentless arms race between bacteria and bacteria-specific viruses

Yet bacteria are not passive bystanders in the face of such an onslaught—quite the contrary. Bacteria possess numerous immune systems to combat viruses at multiple stages during the viral life cycle, which microbiologists have studied for many decades. By the turn of the twenty-first century, the existing paradigm held that, while diverse, these immune systems constituted only a simple innate response to infection. Unlike multicellular vertebrate organisms, which possess innate immune systems together with elaborate adaptive immune systems that can create and store immunological memory, bacteria had no ability to adapt to new threats.

Enter CRISPR, short for Clustered Regularly Interspaced Short Palindromic Repeats. First detected in 1987 in the bacterium *Escherichia coli* (Ishino et al., 1987), CRISPRs—to put it simply—are bizarre, repeating sections of bacterial DNA that can extend thousands of letters in length. While CRISPRs initially seemed like a rare oddity, a fluke of nature, researchers



Color-scanning electron microscope (SEM) image of bacteria on the surface of a human tongue. Enlarged 10,000 times to a width of 10 cm





had detected CRISPRs in dozens of other bacterial species by the early 2000s (Mojica et al., 2000).

These repeating structures were initially described using a number of different and confusing acronyms, and so, in 2002, Dutch researchers simplified their classification with the informative (and catchy) acronym that we still use today (Jansen et al., 2002).

Despite a growing appreciation that CRISPRs were abundant in nature, being found in the genomes of a third of all bacteria and almost all archaea (another domain of single-celled microorganisms), their biological function remained a complete mystery until 2005, when the first clues surfaced linking CRISPR to antiviral immunity (Mojica et al., 2005). Using bioinformatics analyses, researchers were shocked to find viral DNA sequences buried within those repeating sections of DNA, as if the bacteria had somehow stolen the viral genetic code as a form of molecular memory. Might this information allow bacteria to recognize and destroy viral DNA during an infection?

Evidence supporting this hypothesis came from elegant experiments conducted at a yogurt company (Barrangou et al., 2007). Scientists there were hoping to generate virus-resistant strains of the bacterium *Streptococcus thermophilus*, the major workhouse ingredient used to ferment milk into yogurt and other dairy products, and they noticed that, like *E. coli*, their *S. thermophilus* strains also contained CRISPRs. By intentionally infecting their strains with a panel of different viruses and then analyzing the DNA of those bacteria that gained immunity, the researchers proved that CRISPRs indeed conferred adaptive immunity. Almost overnight, the long-standing presumption that bacteria and archaea possessed only comparatively simple defenses against viral pathogens was overturned. Instead, these simple microorganisms employed both innate and adaptive immune systems no less remarkable and versatile than the innate and adaptive systems found in multicellular organisms.

## **Despite a growing appreciation that CRISPRs were abundant in nature, being found in the genomes of a third of all bacteria and almost all archaea (another domain of single-celled microorganisms), their biological function remained a complete mystery until 2005, when the first clues surfaced linking CRISPR to antiviral immunity**

After this breakthrough, it was up to geneticists and biochemists to determine how CRISPR immune systems work. Namely, what enzymes were involved, and how were they able to accurately recognize unique features of viral DNA during an infection? From the work of countless researchers all around the world, a new, unified understanding began to emerge: bacteria and archaea used molecules of ribonucleic acid, or RNA—DNA's molecular cousin—to identify matching sequences of DNA in the viral genome, along with one or more proteins encoded by CRISPR-associated genes to slice apart the DNA (Klompe & Sternberg, 2018). CRISPR was nothing more than a precision-guided pair of molecular scissors, with the incredible ability to home in on specific sequences of DNA and neutralize them by severing both strands of the double helix. And the star actor in this pathway was a protein enzyme called CRISPR-Cas9 (Gasiunas et al., 2012; Jinek et al., 2012).





Resolving the molecular function of CRISPR-Cas9 not only helped solve a key question in bacterial antiviral immune systems. It also immediately revealed immense potential to disrupt a different and seemingly unrelated area of biotechnology: gene editing (Urnov, 2018).

Gene editing refers to a technique in which DNA sequences are modified, or “edited,” directly in the genome of living cells. While effective tools for gene editing in bacteria have been available for decades, the ability to edit DNA in eukaryotic cells, which house the genome in a separate structure called the nucleus, lagged far behind. But in the 1990s, a new strategy for high-efficiency gene editing emerged: if a specific DNA break could be induced at the gene of interest, then the ability to edit that gene was vastly enhanced (Rouet et al., 1994). Somewhat paradoxically, localized DNA damage could serve as a stimulus for DNA repair.

Why might this be the case? Our cells suffer DNA damage constantly, whether from carcinogens or exposure to ionizing radiation, and they have therefore evolved mechanisms for repairing DNA lesions. Detection of a DNA break leads to recruitment of endogenous enzymes to perform this repair, and, over the years, researchers realized that this natural process could be hijacked to install user-defined edits during the repair process. The bottleneck for realizing the full potential of this approach, then, was developing tools to introduce DNA breaks at specific sites in the genome.

The ideal tool would be a “programmable nuclease”—an enzyme that cuts nucleic acids like DNA (hence, “nuclease”), which scientists could easily and rapidly program to recognize and introduce breaks in specific DNA sequences inside the cell (Chandrasegaran & Carroll, 2016). The first such tools were developed in the 1990s and early 2000s, but they were unwieldy, unreliable, and expensive. A researcher might devote months to building a programmable nuclease himself/herself, or spend tens of thousands of dollars outsourcing the work to a company, only to find out that the tool barely worked. In short, gene editing, though validated as a technology, could not realize its full potential because programmable nucleases were simply too hard to engineer.

The discovery of CRISPR-Cas9 offered the perfect solution. Instead of trying to reinvent the wheel, why not harness the programmable nucleases that nature had already sculpted over billions of years of evolution? Whereas bacteria were employing CRISPR-Cas9 to introduce DNA breaks in viral genomes, to prevent infection, perhaps scientists could employ CRISPR-Cas9 to introduce DNA breaks in eukaryotic genomes, to edit genes. The very same property that made CRISPR-Cas9 so effective in adaptive immunity—its ability to precisely home in on DNA targets using an RNA “guide”—might transform researchers’ ability to program nucleases to break specific DNA targets and mark them for repair.

### The CRISPR Craze Begins

In June 2012, Doudna, Charpentier, and colleagues published the first article describing CRISPR-Cas9’s essential components and detailing its utility for gene editing (Jinek et al., 2012). Six months later, the first reports surfaced, demonstrating the remarkable effectiveness of CRISPR-Cas9 for gene editing in both mouse and human cells (Cong et al., 2013; Mali et al., 2013). Within months of that, the first gene-edited mice were created with CRISPR, followed in quick succession by proof-of-concept experiments in rice, rats, wheat, and monkeys, and an increasingly dizzying array of other plant and animal model organisms. The “CRISPR craze” was underway (Pennisi, 2013).





Along with a surge in the species whose genomes could now be seamlessly tweaked with CRISPR, 2013 also witnessed an explosion in the kinds of DNA changes that could be accomplished with CRISPR technology. Beyond fixing small typos in the genome, such as the kinds of mutations that cause genetic disease, CRISPR could be leveraged to inactivate or delete entire genes, invert or insert genes, and make changes to multiple genes simultaneously. An entirely distinct category of applications involved the use of a non-cutting version of CRISPR-Cas9, in which the goal was to ferry other payloads to specific genes in order to turn genes on or off, up or down. By altering gene expression without changing the actual sequence of DNA, researchers could begin to control the very same molecular cues that instructed cells to turn into the many different tissues in the body, all using the same underlying genetic code.

## **2013 witnessed an explosion in the kinds of DNA changes that could be accomplished with CRISPR technology. Beyond fixing small typos in the genome, such as the kinds of mutations that cause genetic disease, CRISPR could be leveraged to inactivate or delete entire genes, invert or insert genes, and make changes to multiple genes simultaneously**

Technology development quickly expanded the CRISPR toolkit, which attracted more and more researchers to the budding gene-editing field. Even more important for CRISPR's widespread adoption than its sheer usefulness, though, was the lowered barrier to entry for novices. First, the ease of engineering CRISPR to target new sites in the genome meant that scientists with a basic understanding of molecular biology could now access what was once an advanced technology requiring years of expertise. (Indeed, some middle- and high-school students now perform CRISPR gene-editing experiments in the classroom [Yu, 2017].) Second, the necessary reagents to perform gene editing could be affordably purchased from nonprofit organizations like Addgene, which distributes CRISPR tools to academic researchers for just \$60 (Kamens, 2015). The result has been a swift, worldwide spread of the technology.

Today, CRISPR encompasses a core set of techniques that biomedical scientists must be well versed in, regardless of their particular research focus, model organism, or preexisting skill set. The technology has quickly become indispensable for performing cutting-edge research, and it is safe to say that biology will never be the same again.

Nor will society. Indeed, armed with the power to easily and precisely rewrite genetic code, scientists and nonscientists alike threaten to upend the natural order, reshaping the very process of evolution by substituting random mutation—the aimless, meandering process acted on by natural selection over the eons—with user-defined mutation, introduced at will via CRISPR technology. The result: a newfound mastery over the direction of life itself.

### **Imminent Impacts on the Planet's Plants and Animals**

Imagine a future world in which you could clone your deceased dog, while also installing DNA mutations that confer hyper-musculature; or in which you could grow super-strains of tomatoes that maintain ripeness long after being picked, mushrooms that do not brown during prolonged storage, and grape vines that are impervious to fungal pests. Out in the



**Gene-edited products have been engineered in the lab, but unlike GMOs, these products do not carry a shred of foreign DNA in the genome, nor any scar from the scientific intervention**

A tray of genetically modified corn grains at Monsanto's stand at Farm Progress, the most important outdoor agriculture fair in the United States. Illinois, August 2017



**CRISPR technology has become indispensable for performing cutting-edge research, and it is safe to say that biology will never be the same again**

The first pig cloned in China, stuffed and exhibited at the Beijing Genomics Institute in Shenzhen, July 2010. According to the World Bank, China's population with an increase from 1,330 million in 2009 to 1,440 million by 2030 has led Beijing to seek cutting-edge technology to help feed its inhabitants and improve food quality



countryside, farmers' pastures accommodate new breeds of dairy cattle, which still retain the same prized genetics resulting from hundreds of years of selective breeding, but no longer develop horns, thanks to gene editing. The nearby pigs possess special mutations that confer viral resistance and also cause them to develop leaner muscles with reduced fat content. In the medical facility one town over, other pigs harbor "humanized" genomes that have been selectively engineered so that the pigs might one day serve as organ donors for humans. Believe it or not, every one of these seemingly fictitious inventions has already been accomplished with the help of CRISPR technology, and the list could go on and on (Doudna & Sternberg, 2017).

Plant breeders are excited by the possibility of engineering new traits into major cash crops with a method that is both safer and more effective than the random mutagenesis methods of the mid- to late-twentieth century, and less invasive than the techniques commonly used to create genetically modified organisms, or GMOs. GMOs are the product of gene splicing, whereby foreign DNA sequences are forcibly integrated into the genetic material of the organism being modified. While no credible evidence exists suggesting that GMOs are any less safe than unmodified plants, they remain the subject of intense public scrutiny and vociferous criticism.

## **Time will tell whether activist groups or overly restrictive regulation will stunt innovation in this sector. One thing seems clear: different cultures and attitudes in distinct parts of the world will play a major role in shaping the future direction of CRISPR applications**

So how will the public then react to gene-edited organisms, which could hit the supermarket in just years (Bunge & Marcus, 2018)? Like GMOs, these products have been engineered in the lab, with the goal of achieving improved yield, enhanced resistance to pests, better taste, or healthier nutritional profile. Unlike GMOs, though, these products do not carry a shred of foreign DNA in the genome, nor any scar from the scientific intervention. In fact, the surgically introduced gene edits are often no different than the DNA mutations that could have arisen by chance. Should we view plant foods any differently if humans introduced a certain mutation, rather than "natural" causes? In spite of the often strident protest against biotechnology products that end up on the dinner table, there are defensible reasons to aggressively pursue gene editing in agricultural improvement if these efforts could address global hunger, nutritional deficiencies, or farming challenges provoked by the future effects of a changing climate.

Time will tell whether activist groups or overly restrictive regulation will stunt innovation in this sector. One thing seems clear: different cultures and attitudes in distinct parts of the world will play a major role in shaping the future direction of CRISPR applications. In the United States, for example, the Department of Agriculture decided in 2018 that plants developed through gene editing will not be specially regulated, as long as they could have been developed through traditional breeding. In stark contrast, the European Court of Justice decided around the same time that gene-edited crops would be subject to the same regulations as GMOs. Meanwhile, the application of CRISPR in agriculture surges ahead in China, which ranks first in worldwide farm output.







Food producers are equally excited by the possibilities afforded by gene-editing technology in animals. Designer DNA mutations can increase muscle content, reduce disease, and make animals, and they also offer biological solutions to problems often solved through more ruthless means. For example, rather than farmers killing off male chicks a day after hatching because female hens are desired, scientists are pursuing gene-editing solutions to bias reproduction, so that only female chicks are born in the first place. Similarly, the remarkable feat by a company called Recombinetics to genetically “dehorn” cattle offers a far more humane alternative to the cruel but widespread practice of cattle dehorning via cauterization. Gene editing could even succeed in producing chickens that lay hypoallergenic eggs, pigs that do not require as many antibiotic treatments, and sheep with naturally altered hair color.

Then there are those applications of CRISPR that verge on science fiction. Rather than harnessing gene-editing technology to create organisms never before seen on Earth, some scientists aim to do exactly the opposite and leverage gene editing to resurrect extinct animals that once existed long ago. Dinosaurs are sadly out of the question, as imagined by Michael Crichton in *Jurassic Park*—DNA breaks down far too quickly to rebuild the genome of any dinosaur species—but not so with the woolly mammoth. Using extremely well-preserved frozen tissue samples, geneticists have already succeeded in deciphering the letter-by-letter sequence of the woolly mammoth genome, enabling a direct comparison to the genome of the modern-day elephant, its closest relative. Now, George Church and colleagues are using CRISPR to convert specific genes in elephant cells into their woolly mammoth counterparts, prioritizing those genes implicated in functions like temperature sensation, fat tissue production, and skin and hair development. Organizations like the Long Now Foundation hope to bring genetic engineering to bear on many more such de-extinction efforts, with a focus on passenger pigeons, great auks, and gastric-brooding frogs, all of which were directly or indirectly wiped off the planet by human actions. Might we be able to reverse some of the negative impacts that humans have had on biodiversity, using biotechnology? Or should we, instead, be focusing our efforts on preserving the biodiversity that is left, by working harder to curb climate change, restrain poaching, and rein in excessive land development?

**There are CRISPR applications that verge on science fiction. Rather than harnessing gene-editing technology to create organisms never before seen on Earth, some scientists aim to do exactly the opposite and leverage gene editing to resurrect extinct animals that once existed long ago**

One additional application of CRISPR in animals deserves mention: a potentially revolutionary technology known as a gene drive (Regalado, 2016). The scientific details are complicated, having to do with a clever workaround of those fundamental laws of inheritance first discovered by Gregor Mendel through his work on pea plants. CRISPR-based gene drives allow bioengineers to break those laws, effectively “driving” new genes into wild animal populations at unprecedented speed, along with their associated traits. Perhaps the most promising real-world example involves the mosquito, which causes more human suffering than any other creature on Earth because of its extraordinary ability to serve as a vector for countless viruses (dengue, West Nile, Zika, and so on), as well as the malaria parasite. Imagine if genet-

ically modified mosquitoes, specifically engineered so they can no longer transmit malaria, were released into the wild and allowed to spread their genes. Better yet, what about a line of genetically modified mosquitoes that spread female sterility, thereby hindering reproduction and culling entire wild populations? Proof-of-concept experiments achieving both these feats have already been performed in highly contained laboratory environments, and discussions are underway to determine when the technology is safe enough for field trials. Attempting the eradication of an entire species may rightfully seem like a dangerously blunt instrument to wield, and yet, if mosquito-borne illnesses were to become a thing of the past, saving a million human lives annually, can we justify *not* taking the risk?



### Realizing the Promise of Gene Editing to Treat Disease

Notwithstanding the numerous exciting developments in plant and animal applications, the greatest promise of CRISPR technology is arguably to cure genetic diseases in human patients (Stockton, 2017). Monogenic genetic diseases result from one or more mutations in a single gene, and scientists estimate that there are more than 10,000 such diseases affecting roughly one in every two hundred births. Many genetic diseases like Tay-Sachs disease are fatal at a young age; others like cystic fibrosis can be managed but still lead to a significant reduction in life expectancy; and still others lead to devastating outcomes later in life, such as the physical, mental, and behavioral decline that inevitably occurs for Huntington's disease patients.

## Notwithstanding the numerous exciting developments in plant and animal applications, the greatest promise of CRISPR technology is arguably to cure genetic diseases in human patients

Scientists have been dreaming of a magic bullet cure for genetic diseases ever since DNA mutations were first linked to hereditary illnesses, and there have been incredible strides over the years. For example, after more than twenty-five years of clinical trials, the first gene therapy drug was approved by the United States Food and Drug Administration in 2017, in which patients suffering from a disease of the eye called retinal dystrophy receive healthy genes delivered directly into the eye via a genetically modified virus. Other diseases can be effectively treated using small-molecule drugs, or, in more severe cases, bone marrow transplants. Yet all of these approaches treat the genetic disease indirectly, rather than directly targeting the causative DNA mutations. The ideal treatment would permanently cure the disease by repairing the mutation itself, editing the pathogenic DNA sequence back to its healthy counterpart.

CRISPR offers the possibility of this ideal treatment. In dozens of proof-of-concept studies already published, scientists have successfully leveraged CRISPR in cultured human cells to eradicate the mutations that cause sickle cell disease, beta-thalassemia, hemophilia, Duchenne muscular dystrophy, blindness, and countless other genetic disorders. CRISPR has been injected into mouse and canine models of human disease, and achieved lasting and effective reversal of disease symptoms. And physicians have already tested the first gene-editing-based treatments in patients, though it is too early to say whether or not the treatments were efficacious.



A researcher initiates a CRISPR-Cas9 process at the Max-Delbrueck Molecular Medicine Center in Berlin, Germany.  
May 22, 2018



In a parallel and equally exciting avenue of research, CRISPR is being combined with a promising (and Nobel Prize-winning) new avenue of cancer treatment, known as cancer immunotherapy. Here, human immune cells are enhanced with genetic engineering, endowing them with specialized molecules that can hunt down markers specific to cancer, and then potentially eliminate cancerous cells from the body. In a remarkable first, Layla Richards, a one-year-old patient from London who was suffering from acute lymphoblastic leukemia, the most common type of childhood cancer, was cured in 2015 using a combination of gene-edited immune cells and bone marrow transplant. Chinese scientists have since initiated clinical trials in dozens of other patients using gene-edited immune cells to treat cancer, and additional trials are imminent in the US and Europe.

To be sure, many challenges remain before the full potential of CRISPR-based disease cures can be realized. For one, the tricky problem of delivery remains: how to deliver CRISPR into the body and edit enough of an adult patient's forty trillion cells to have a lasting effect, and to do so safely without any adverse effects. Additionally, the edits need to be introduced with an extreme level of accuracy, so that other genes are not inadvertently perturbed while the disease-associated mutation is being repaired. Early reports highlighted the risk of so-called off-target effects, in which CRISPR induced unintended mutations, and, given the permanent nature of DNA changes, the bar must be set very high for a gene-editing therapy to be proven safe.

**CRISPR is being combined with a new avenue of cancer treatment, known as cancer immunotherapy, in which human immune cells are enhanced with genetic engineering, endowing them with specialized molecules that can hunt down markers specific to cancer, and then potentially eliminate cancerous cells from the body**

In spite of the risks and remaining hurdles, the possibilities seem limitless. New studies now surface at a rate of more than five per day, on average, and investors have poured billions of dollars into the various companies that are now pursuing CRISPR-based therapeutics. Someday soon, we may find ourselves in a new era in which genetic diseases and cancer are no longer incurable maladies to be endured, but tractable medical problems to be solved.

#### The Looming Ethical Controversy over Embryo Editing

When should that solution begin, though? While most researchers are focused on harnessing CRISPR to treat patients living with disease, a small but growing number of scientists are, instead, exploring the use of CRISPR to *prevent* disease, not in living patients but in future individuals. By repairing DNA mutations directly in human embryos conceived in the laboratory, from the merger of egg and sperm through *in vitro* fertilization (IVF), these scientists hope to create heritable gene edits that would be copied into every cell of the developing human and passed on to all future offspring.

Introducing gene edits into embryos constitutes a form of germline editing, in which the germline refers to any germ cells whose genome can be inherited by subsequent generations. Germline editing is routinely practiced by animal breeders because it is the







most effective way of creating genetically modified animals, and, indeed, methods for injecting CRISPR into mouse embryos have been all but perfected over the last five years. Yet the notion of performing similar experiments in human embryos is cause for alarm, not just because of heightened safety concerns when introducing heritable mutations, but because of the ethical and societal ramifications of a technology that could forever alter the human genome for generations to come.

In 2015, when it became clear that CRISPR would make human germline editing a distinct possibility, numerous white papers were published by scientists and nonscientists alike, calling attention to this troubling area of technology development. In almost perfect synchrony, though, the first research article was published by a group of Chinese scientists in which, for the first time ever, human embryos were subjected to precision gene editing. The resulting embryos were not implanted to establish pregnancies, and those initial experiments were not particularly successful at achieving the desired edits, but, nevertheless, the red line had been crossed. In the years since, additional studies have continued to surface, and in one of the most recent articles from a group in the US, the technique was shown to be far safer than before, and far more effective. Many fear that the first humans to be born with engineered DNA mutations may be just around the corner.

The media is rife with doomsday scenarios auguring a future of designer babies with superhuman intelligence, strength, or beauty, and it is important to realize the flaws in such alarmist reactions. The vast majority of human traits can only partially be ascribed to genetics, and they typically result from thousands and thousands of gene variants, each one of which has only a vanishingly small impact on determining the trait. It is hard enough to introduce a single mutation precisely with CRISPR, and no amount of gene editing would be able to achieve the thousands of edits required for attempting to alter these traits. The sci-fi futuristic scenarios depicted by movies like *Gattaca* and books like *A Brave New World* are bound to remain just that: science fiction.

**In 2015, for the first time ever, human embryos were subjected to precision gene editing. The resulting embryos were not implanted to establish pregnancies, and those initial experiments were not particularly successful at achieving the desired edits, but, nevertheless, the red line had been crossed**

Nevertheless, the emergence of facile, powerful, gene-editing technology may change the way we think about reproduction, particularly when it comes to highly penetrant, disease-associated mutations. If it becomes possible to eradicate a mutation before birth, eliminating the chance that a child could ever develop a particular disease, or pass on the disease-associated mutation to his/her children, should we not pursue such an intervention? But how might such interventions change the way society perceives individuals already living with disease? Who would have access to such interventions, and would they be offered equitably? And might the use of CRISPR to eliminate disease-associated mutations simply be the first step down the slippery slope of harnessing gene editing for genetic enhancements?

These are hard questions, questions that must be discussed and debated, and not just by scientists but by the many other stakeholders who will be affected by gene-editing technology:

patients and patient advocacy groups, bioethicists, philosophers, religious leaders, disability rights advocates, regulators, and members of the public. Furthermore, we must endeavor to reach across cultural divides and seek international consensus, thereby avoiding a potential genetic arms race in which countries compete to innovate faster than others.

There are real risks associated with the unfettered pursuit of human germline editing, but this must not be allowed to stifle the development of CRISPR for improving our society in other ways. Few technologies are inherently good or bad: what is critical is how we use them. The power to control our species' genetic future is both terrifying and awesome, and we must rise to the challenge of deciding how best to harness it.



## Concluding Thoughts

Ten years ago, the term CRISPR was familiar to just a few dozen scientists around the world. Today, hardly a day passes without a feature story touting the amazing possibilities of CRISPR technology, and gene editing is quickly becoming a household term. CRISPR has starred in a Hollywood blockbuster movie, appeared in countless TV series, been discussed by governmental agencies worldwide, and become available online for purchase as a do-it-yourself kit. Ten years hence, CRISPR will impact the food we eat and the medicine we take, and it will undoubtedly continue to prove instrumental in our understanding of the natural world all around us.

So, too, will our understanding of CRISPR itself continue to evolve. For the field of bacterial adaptive immunity is anything but stagnant, and new discoveries abound as researchers continue to dig deeper and deeper into the billion-year-old genetic conflict between bacteria and viruses. We now know that CRISPR-Cas9 is just one of many remarkable molecular machines that microbes have evolved to counteract the perpetual assault from foreign pathogens, and scientists continue to invent innovative applications of this biological treasure trove. Who would have thought that scientific curiosity and basic research investigations could unveil such a promising area of biotechnological exploration?

The American physicist, Leonard Susskind, once wrote: "Unforeseen surprises are the rule in science, not the exception." Let us see where the next big breakthrough comes from.

## Acknowledgments

I thank Abigail Fisher for assistance with early drafts of the article outline. I regret that space constraints prevented me from more thoroughly discussing the research being led by my many colleagues, and I encourage interested readers to consult the included references for a more thorough discussion of related work.

## Select Bibliography

- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., and Moineau, S., et al. 2007. “CRISPR provides acquired resistance against viruses in prokaryotes.” *Science* 315(5819): 1709–1712.
- Bunge, J., and Marcus, A. D. 2018. “Is this tomato engineered? Inside the coming battle over gene-edited food.” *The Wall Street Journal*, 15 April.
- Chandrasegaran, S., and Carroll, D. 2016. “Origins of programmable nucleases for genome engineering.” *Journal of Molecular Biology* 428(5 Pt B): 963–989.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. 2013. “Multiplex genome engineering using CRISPR/Cas systems.” *Science* 339(6121): 819–823.
- Doudna, J. A., and Sternberg, S. H. 2017. *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*. New York: Houghton Mifflin Harcourt.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksny, V. 2012. “Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.” *Proceedings of the National Academy of Sciences of the United States of America* 109(39): E2579–86.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. 1987. “Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product.” *Journal of Bacteriology* 169(12): 5429–5433.
- Jansen, R., Embden, J. D. A. V., Gaastra, W., and Schouls, L. M. 2002. “Identification of genes that are associated with DNA repeats in prokaryotes.” *Molecular Microbiology* 43(6): 1565–1575.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. 2012. “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” *Science* 337(6096): 816–821.
- Kamens, J. 2015. “The Addgene repository: An international nonprofit plasmid and data resource.” *Nucleic Acids Research* 43, D1152–7.
- Keen, E. C. 2015. “A century of phage research: Bacteriophages and the shaping of modern biology.” *BioEssays* 37(1): 6–9.
- Klompe, S. E., and Sternberg, S. H. 2018. “Harnessing ‘a billion years of experimentation’: The ongoing exploration and exploitation of CRISPR-Cas immune systems.” *The CRISPR Journal* 1(2): 141–158.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. 2013. “RNA-guided human genome engineering via Cas9.” *Science* 339(6121): 823–826.
- Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. 2005. “Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.” *Journal of Molecular Evolution* 60(2): 174–182.
- Mojica, F. J., Díez-Villaseñor, C., Soria, E., and Juez, G. 2000. “Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.” *Molecular Microbiology* 36(1): 244–246.
- Pennisi, E. 2013. “The CRISPR craze.” *Science* 341(6148): 833–836.
- Regalado, A. 2016. “The extinction invention.” *MIT Technology Review*, 13 April.
- Rohwer, F., Youle, M., Maughan, H., and Hisakawa, N. 2014. *Life in Our Phage World*. San Diego, CA: Wholon.
- Rouet, P., Smih, F., and Jasin, M. 1994. “Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.” *Proceedings of the National Academy of Sciences of the United States of America* 91(13): 6064–6068.
- Stockton, N. 2017. “How CRISPR could snip away some of humanity’s worst diseases.” *Wired*, 5 May.
- Urnov, F. D. 2018. “Genome editing B.C. (Before CRISPR): Lasting lessons from the ‘Old Testament.’” *The CRISPR Journal*, 1(1): 34–46. <https://doi.org/10.1089/crispr.2018.29007.fyu>.
- Yu, A. 2017. “How a gene editing tool went from labs to a middle-school classroom.” *All Tech Considered*, 27 May.



**Peter Kalmus**  
Climate Scientist

Dr. Peter Kalmus is a climate scientist at NASA's Jet Propulsion Laboratory in Pasadena, California (writing on his own behalf). His research interests include low clouds, tornadoes, ecological forecasting, and improving the utility of satellite climate records. He is the recipient of NASA's Early Career Achievement medal, the author of the award-winning book *Being the Change: Live Well and Spark a Climate Revolution*, and the founder of [noflyclimatesci.org](http://noflyclimatesci.org). Knowing the long-lasting harm that it causes, Peter now loathes burning fossil fuel. He lives an outwardly normal life with his wife and two sons, but on about a tenth the fossil fuel of the average American.

Book recommended: *Being the Change: Live Well and Spark a Climate Revolution*, Peter Kalmus, New Society, 2017.

**Climate change—or perhaps more aptly, climate breakdown—is the greatest challenge facing humanity today. Fossil fuel is entangled in every aspect of modern life, but burning it releases carbon dioxide, an invisible gas that warms the Earth via infrared absorption and remains in the atmosphere for thousands of years. By 2018, warming of about 1.2°C beyond the preindustrial baseline has already caused unacceptable impacts, but these impacts will worsen precipitously as warming proceeds. The previous decade saw progress in climate science, but it also saw a procession of devastating climate-change-related natural disasters affecting both humans and nonhumans. While humanity's sense of urgency is growing, it remains far below the level required to avoid catastrophic warming that would threaten civilization as we know it.**



The next few years are probably the most important in (human) history.  
Dr. Debra Roberts, IPCC Working Group II Co-Chair



It is hard to know where to start when writing about climate change, as it affects nearly every aspect of human life on this planet, and at every scale. It affects our security, our food and water systems, our energy and economic systems, our infrastructure. It affects the intensity and cost of natural disasters. It affects most any ecosystem you care to study, whether terrestrial or marine. It affects our mental health, the priorities of our communities and cities, the futures of our children. It affects our politics, the openness of our societies, how we as nations relate to other nations, and how we as individuals relate to each other, especially those we see as not part of our tribe.

At only 1.2°C of mean global surface warming, many of these aspects of climate change have already become not just visible, but catastrophic. And it is crucial to realize that every climate impact will intensify as warming proceeds. In other words, what we are experiencing today is far from a “new normal”: impacts will get worse, and worse, and worse as warming proceeds. Indeed, the phrase “climate change” may no longer adequately capture the urgency of this reality, which might better be described as “climate breakdown”; and many climate scientists, myself included, are now embracing the label “alarmist.” For there is an alarm to sound.

Climate breakdown is a devilish problem for humanity. During the nineteenth century, a handful of scientists worked out how fossil fuel emissions were warming the planet. These early climate scientists were not alarmed about something so comfortably in the future; some even thought that a warmer Earth would be a good thing. With each passing decade, fossil fuel became more deeply entwined into the economy and into daily life, from transportation to construction to manufacturing to food. The infusion of fossil fuel drove spectacular increases in wealth, mobility, consumption, technology, and medicine. It powered a Green Revolution in agriculture, and the exponential growth in food inevitably drove an exponential growth in population. Human civilization today literally runs on fossil fuel. We are addicted to the stuff: try to imagine a world without it. Try to imagine a military voluntarily giving it up. As individuals, communities, corporations, and nations, we are deeply attached to the convenient, profitable, *powerful* stuff.

**The phrase “climate change” may no longer adequately capture the urgency of this reality, which might better be described as “climate breakdown”; and many climate scientists, myself included, are now embracing the label “alarmist.” For there is an alarm to sound**

But the devil will have its due. By the 1960s, scientists began to realize the dangerous implications of global warming and to sound the alarm. A 1965 White House report, for example, warned that continued fossil fuel use would lead to “apocalyptic” and “irreversible climate change,” including just over three meters (ten feet) of sea-level rise. Despite these clear warnings, the US government responded by doing nothing. The fossil fuel corporations at first made good-faith attempts to understand the problem. But in the 1980s, sensing imminent climate action that could threaten their massive profits, they literally chose to sell out the



entire world, embarking on a systematic misinformation campaign to sow confusion among the general public and delay action for as long as possible. Meanwhile, the public lacked even the usual, visible indications of pollution, such as garbage lying on the ground or smog in the air. Burning fossil fuel emits carbon dioxide (CO<sub>2</sub>), a gas that naturally occurs in the atmosphere and traps outgoing infrared radiation, warming the Earth. Because it is odorless and colorless, CO<sub>2</sub> emissions fly below the radar of our awareness. And due to its incremental and global nature, it is easy to think of climate change as something safely far away or in the future. Even environmental activists were largely unaware of the climate threat until at least the 1980s.

This lack of immediacy is perhaps the biggest block to climate action, as the public continues to blithely ignore climate change at the voting booth, placing it at the bottom of their list of issues. However, even as our lives proceed apace and our day-to-day experience seems normal—although there are hints for the observant, even in our own backyards, of the massive changes underway—there is now a growing consensus among Earth scientists that climate change poses an existential threat to human civilization.

In other words, civilization is at a crossroads. Humans have become the dominant perturbation to the natural world at a global scale, mainly via two mechanisms: by expanding into and transforming the habitats of other species, and by emitting greenhouse gases and warming the planet.

## **Lack of immediacy is perhaps the biggest block to climate action, as the public continues to blithely ignore climate change at the voting booth, placing it at the bottom of their list of issues**

Today, humanity must choose to transition its entire energy infrastructure away from fossil fuel in an incredibly short span of time—within the next thirty years—or suffer increasingly catastrophic consequences.

This brief overview of the potential of climate breakdown to disrupt humanity's progress is no doubt a sobering counterweight to the exuberance on display elsewhere in this volume. While by some metrics humanity is doing better than ever, the overall picture is somewhat bleaker, as climate breakdown has the potential to derail all this progress. To minimize this risk, humanity must address it with the utmost urgency, with an energy of mobilization and level of cooperation never before seen in the history of our species. And unfortunately, humanity as a whole is so far showing little of this necessary urgency; indeed, in recent years there has even been some regress, as populist, anti-science regimes gain power in the United States, Europe, and elsewhere. The challenge we face as a species is indeed sobering. While those who claim that humanity is doing better than ever are bound to be the more popular messengers, perhaps this is not the time to party, but instead to get to work.

The rest of this article is organized as follows. I first look retrospectively at the decade before 2018, sampling a few climate milestones. I then discuss a few of the climate impacts projected to occur at just 1.5°C of warming and 2°C of warming. Next, I discuss some potential solutions. I argue that we need myriad solutions at all scales, an “all of the above” approach, but that we must avoid attractive but dangerous pseudo-solutions that may not turn out to be solutions at all. Finally, I zoom far out to consider our climate conundrum from an astronomical perspective, concluding with thoughts about what a “new enlightenment” might mean in the context of climate breakdown.

Because writing such a broad overview requires me both to wander outside my areas of scientific expertise and to offer my opinions, I write this paper primarily from my role as a citizen.



## The Last Ten Years in Climate Change

The decade prior to 2018 saw increasing greenhouse gas emissions and warming, advances in climate science, an increase in climate-change-related disasters and costs, and an increasing sense of urgency and action.

Perhaps a good way to kick off a brief (and necessarily incomplete) overview of the last ten years in climate change is with the following simple but astounding facts: Antarctica is melting three times faster now than it was just a decade ago.<sup>1</sup> This is the nature of ice in a warming world.

## **Warming over the period 2008 to 2017 is clearly apparent in a variety of independent heat indicators. That these signals can emerge from Earth system variability over such a short period is a remarkable testament to the unprecedented rapidity of warming**

The principal driver of warming is human CO<sub>2</sub> emissions from fossil fuel burning and deforestation. In June 2018, the atmospheric CO<sub>2</sub> fraction measured from Mauna Loa in Hawaii was 411 ppm, up from 388 ppm in June 2008.<sup>2</sup> It has been increasing at a steady exponential rate of 2.2% per year since about 1790 and a preindustrial level of 280 ppm.<sup>3</sup> Recent year-over-year increases have averaged about 2 ppm per year. This stubborn exponential increase in atmospheric CO<sub>2</sub> is indicative of just how foundational fossil fuel has been to human civilization since the Industrial Revolution; stopping that growth in a controlled way (as opposed to collapse) will require an energy revolution. Humanity, for its entire existence, has obtained energy primarily by burning things. Our task now is to switch directly to the Sun.

## A Warming World

Warming over the period 2008 to 2017 is clearly apparent in a variety of independent heat indicators. That these signals can emerge from Earth system variability over such a short period is a remarkable testament to the unprecedented rapidity of warming. At no time in Earth's history has it warmed this rapidly.

Figure 1 shows the global mean surface temperature since 1850, relative to the mean value of 1850–1900, from the Berkeley Earth dataset which blends sea surface temperatures over ocean and near-surface air temperatures over land.<sup>4</sup> The 2015–17 mean value was 1.2°C above the baseline; the 2017 value estimated from a thirty-one-year extrapolated mean is 1.15°C above the baseline.<sup>5</sup> The global mean surface warming has recently increased by 0.20°C per decade (as determined from a linear fit from 1988 to 2017). The recent decade included seven of the ten hottest years on record. If global temperature continues rising at this rate, it implies attainment of 1.5°C above the baseline in about 2035, and attainment of 2°C above the baseline in about 2060.



Bathers at Wannsee, near Berlin, in July 2014, when temperatures reached over 30°C in Germany. The customary temperature for that time of year is 20°C. The year 2014 was the hottest in recorded history





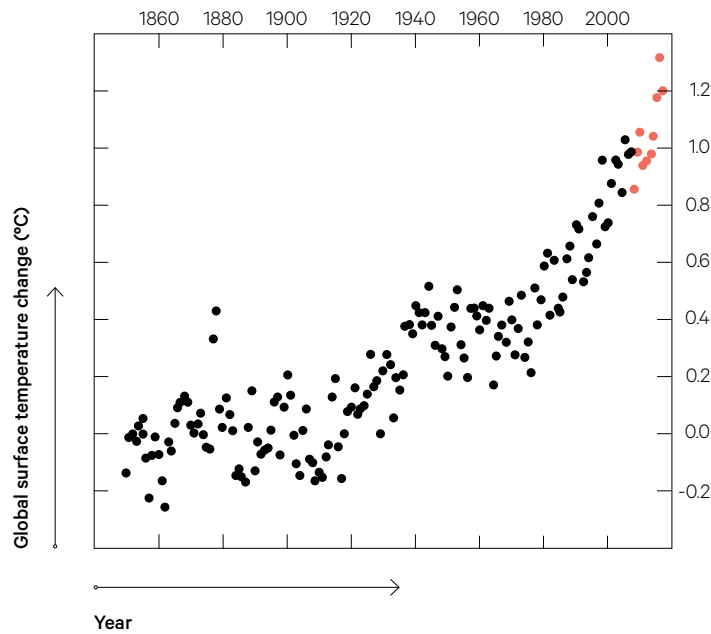


Figure 1. Global average surface temperature relative to the 1850–1900 mean through 2017, from Berkeley Earth. The decade 2008–17 is shown in red

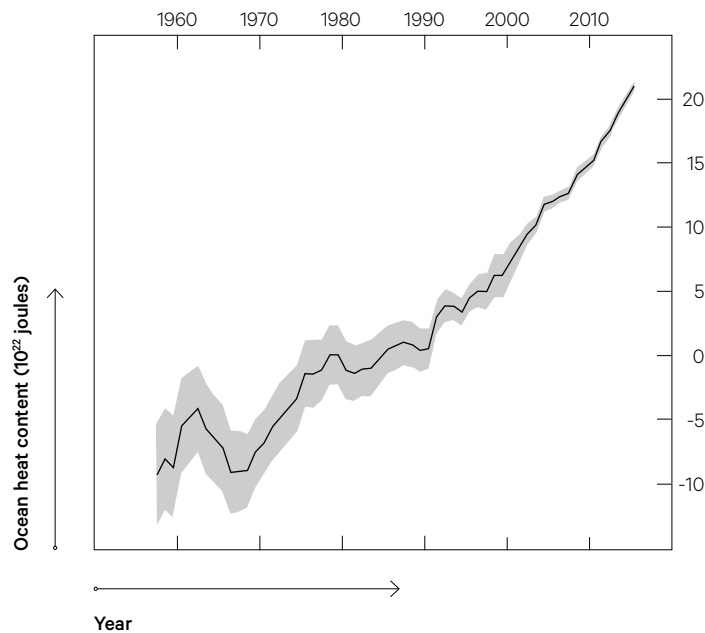


Figure 2. Global average ocean heat content relative to the 1955–2006 mean. Shading gives the 90% confidence interval



Ocean heat content is an even more reliable indicator of global warming, as over 90% of the Earth's current heat imbalance is pouring into the oceans. Figure 2 shows ocean heat content since 1955 in the layer from the surface down to a depth of two kilometers.<sup>6</sup> Both global mean surface temperature and ocean heat content are increasing so rapidly that it is now becoming pointless to say, for example, that 2017 broke records for ocean heat.

Summer Arctic sea ice volume fluctuates significantly from one year to the next due to a variety of factors; it is connected to the rest of the Earth system. However, it decreased by over 20% from a mean of 18,900 km<sup>3</sup> over the decade 1998–2007 to a mean of 14,700 km<sup>3</sup> over the decade 2008–17.<sup>7</sup> 2017 saw record low Arctic sea ice volume of 12,900 km<sup>3</sup>. In 2017, a tanker carrying natural gas became the first merchant ship to cross the Arctic without an ice breaker; and the summer of 2018 saw the first container ship in history successfully traverse the northern Arctic route.<sup>8</sup> In February 2018, there was a massive, long-lived warming event in the Arctic, with average daily temperatures up to 20°C above the 1958–2002 means. While anomalous and alarming, the implications are not yet clear.<sup>9</sup>

Meanwhile, sea-level rise, which has been monitored from space since 1993, has increased by 3–4 cm over the recent decade. Roughly two thirds of the rise is due to meltwater from ice sheets and glaciers and roughly one third is due to thermal expansion as the ocean warms. In 2017, global mean sea level was 77 mm higher than it was in 1993. It is accelerating by 0.084 mm per year. Sea-level rise varies geographically, with anomalies of up to 15 cm above and below this mean rise.<sup>10</sup> Over the last decade, it has had a profound effect on tropical cyclone storm surges, low-lying nations, and sunny-day flooding in coastal cities, and costs are mounting. Coral reefs were devastated by ocean heat waves between 2014 and 2017. More than 75% of Earth's tropical reefs experienced bleaching stress during this period, and 30% experienced mortality stress levels.<sup>11</sup> Mass bleaching now returns on average every six years, faster than reefs can recover. This is projected to increase as warming progresses.

## **In 2017, a tanker carrying natural gas became the first merchant ship to cross the Arctic without an ice breaker; and the summer of 2018 saw the first container ship in history successfully traverse the northern Arctic route**

There have been a large number of other profound global and regional Earth system changes over the last decade due to global warming: mountain glaciers have melted; extreme heat days have increased; spring has moved earlier; and drought, wildfire, and associated ecological transformation has increased in the US west and elsewhere.

### **Climate-Related Natural Disasters**

The last ten years have seen a rise in climate-related natural disasters and associated costs. Classes of natural disasters with increasing intensity clearly related to global warming include tropical cyclones, wildfires, drought, and flooding.

Hurricanes and typhoons are becoming more frequent and more intense due to rapidly warming oceans, warmer atmospheres that hold more moisture, sea-level rise that intensifies storm surge, and slower motions due to climate-related jet stream changes. Cyclones are now



tending to intensify more rapidly as well. Five of the six costliest Atlantic hurricanes have occurred in the decade 2008–17, the sixth being Hurricane Katrina in 2005.

As many regions become hotter and drier due to climate change, wildfires are worsening. In the state of California, for example, fifteen of the twenty largest fires in history have burned since 2000.<sup>12</sup> 2017 was the most destructive year for wildfires in California on record, and 2018 is on track to break that record.

## **Wild fires, drought, flooding and tropical cyclones are related to global warming. Five of the six costliest Atlantic hurricanes have occurred in the decade 2008–17, the sixth being Hurricane Katrina in 2005**

While drought is challenging to define and measure and has multiple contributing factors, it has recently become clear (since 2013) that global warming is intensifying drought in some regions of the world. Climate-intensified drought can be caused by climate-induced shortfalls in precipitation as well as by higher temperatures, which evaporate soil moisture and decrease snowpack via earlier melt and a shift from snowfall to rain. Climate change is now thought to have intensified recent drought in California and the US southwest, and to have contributed to drought thought to be one factor in precipitating the Syrian civil war in 2011.<sup>13</sup>

The other side of drought is excess rainfall, which can cause flooding and mudslides. Indeed, large-scale teleconnections in the atmosphere (stationary waves) can interconnect both drying in the western US and flooding in South Asia,<sup>14</sup> such as the 2017 monsoon which affected more than forty-five million people, killing over one thousand. In addition to changes in atmospheric dynamics, a warmer atmosphere holds more water, leading to heavier precipitation events.

The insurance industry is facing losses due to these enhanced risks, and has not had time to adapt. 2017 saw record losses in the industry.<sup>15</sup>

### **Advances in Climate Science**

By 2007, climate scientists had long-since provided the world with unequivocal evidence for the most important societal response: burning fossil fuel causes warming, which is bringing catastrophic effects; so stop burning fossil fuel. In that sense, the last decade has seen no breakthroughs in climate science. Still, the scientific community has filled in many important details over this period; here are a handful of my personal favorites.

The use of climate models to probabilistically attribute individual weather events to climate change is advancing rapidly, something thought to be impossible in the early 2000s. In 2004, Stott, Stone, and Allen published the first attribution study for the 2003 deadly heatwave in Europe.<sup>16</sup> Attribution studies are now performed for a wide variety of events. In 2013, the Intergovernmental Panel on Climate Change (IPCC) stated that attribution of individual droughts to climate change was not possible; and even a few years ago journalists would routinely state that no individual event can be attributed to climate change. Today, however, rigorous model-based attributions sometimes appear days after natural disasters,

and soon such attributions will be made in real time.<sup>17</sup> Real-time attribution could potentially help increase the public's awareness of urgency. In general, attribution could have legal implications for corporate carbon polluters.

Remote sensing from satellites has advanced over the last decade. GOSAT (launched 2009) and OCO-2 (launched 2014) provide precision measurements of CO<sub>2</sub> concentrations for the entire planet, a crucial measurement for international cooperation on mitigation. Data from these carbon-monitoring satellites is also crucial for improving our understanding of the global carbon cycle. There has also been a quiet revolution in the remote sensing of ecological systems; for example, GOSAT and OCO-2 also serendipitously provide measurements of solar-induced fluorescence, which allows researchers to deduce plant health, stress, and productivity. In general, for many space-borne measurements which began in the 1990s, data records now extend for more than two decades, making them increasingly useful in climatological contexts.



## **The use of climate models to probabilistically attribute individual weather events to climate change is advancing rapidly, something thought to be impossible in the early 2000s. Stott, Stone, and Allen published the first attribution study for the 2003 deadly heatwave in Europe**

Whereas *in situ* measurements of the atmosphere have long relied mainly on mature technologies, such as weather balloons, radar, and LIDAR, the last decade has seen a revolution in *in situ* ocean measurements with the advent of Argo floats. Argo is a system of about 4,000 floats distributed over the global ocean. The floats measure temperature, salinity, and current, spending most of their time drifting at a depth of one kilometer. Every ten days they descend to two kilometers and then ascend to the surface, where they relay their measurements to satellites. Argo is an international collaboration that went operational with 3,000 floats in 2007, and since then has revolutionized our ability to measure the Earth's energy imbalance.

Climate models have also steadily improved. In 2007, the IPCC's AR4 was released and global climate models (from the third Coupled Model Intercomparison Project, CMIP3) had typical horizontal resolutions of about 110 kilometers; resolution has since improved, and adaptive mesh refinement places high resolution where it is most needed. In 2008, models were used for the first time to study climate tipping points, such as Arctic sea ice and the Greenland ice sheet. Modeling of aerosols were improved, and black carbon is given more attention.<sup>18</sup> CMIP5 served as a foundation for the IPCC AR5, which was released in 2013 and included an in-depth evaluation of the CMIP5 model ensemble. Today, as we head into CMIP6 and IPCC AR6, most projections of global change will come from Earth System Models (ESMs), which include coupled ecosystem and biosphere models in addition to atmosphere, ocean, ice, and land. The trend is toward increasing spatial resolution, as computer equipment continues to become more powerful. Advances in regional modeling and downscaling have also allowed for increasingly useful regional projections.

In summary, we can now monitor, measure, and model the Earth system with more precision and accuracy than ever. However, there is still much room for improvement. For example, the estimated range of equilibrium climate sensitivity to a doubling of CO<sub>2</sub> has been



essentially unchanged at 1.5°C to 4.5°C since the Charney report in 1979. Difficult modeling challenges remain, such as improving the representation of clouds and aerosols, carbon cycle feedbacks, and vegetation; and capturing nonlinearities' tipping points. All models are wrong, but some are useful.<sup>19</sup> The modeling and observational communities have been working together to make models—the bridge to knowledge about the future—more useful.



## **In 2008, models were used for the first time to study climate tipping points, such as Arctic sea ice and the Greenland ice sheet. Modeling of aerosols were improved, and black carbon is given more attention**

Finally, it is worth mentioning that over the last decade it has been found that essentially all working climate scientists (at least 97%) agree that humans are warming the planet.<sup>20</sup> But what about those 3% who deny this consensus? They amounted to thirty-eight peer-reviewed papers over the last decade. It turns out that every single one of them had errors, and when the errors were corrected, the revised results agreed with the consensus in every case.<sup>21</sup>

### **Humanity's Response**

The most important metric for assessing humanity's response over the last decade is CO<sub>2</sub> emissions, and, globally, emissions are still increasing. The unavoidable conclusion is that whatever humanity is doing, it is not working; at least not yet.

In 2017, the four largest emitters were China (28%), the US (15%), the EU (10%), and India (7%).<sup>22</sup> China's emissions began increasing dramatically around 2000 and surpassed the US in 2005. However, China's emissions decreased by 0.3% from 2015 to 2017, while India's emissions increased 4.5% over the same period. India and China have the same population (1.3 and 1.4 billion people, respectively). While India has had a relatively minor part in warming over the recent decade, it will become increasingly important going forward.

In 2013, the nations of the world met in Paris to discuss climate action under the United Nations Framework Convention on Climate Change. Given how little had been accomplished previously on the international stage, even orchestrating this meeting was a significant achievement. If the Paris Agreement were to be honored, it would lead to global greenhouse gas emissions (CO<sub>2</sub> and non-CO<sub>2</sub>) in 2030 of about 55 GtCO<sub>2</sub>-equivalents per year, which is not nearly enough mitigation to keep warming under 1.5°C and would lead to warming well in excess of 3°C.<sup>23</sup> Essentially all available 1.5°C pathways have emissions below 35 GtCO<sub>2</sub>eq/yr in 2030, and most require even less. The plan was to strengthen the Paris Agreement going forward in order to meet targets. However, the United States and President Trump loudly proclaimed that it will not only ignore the agreement, but will actively pursue policies to increase emissions, such as coddling the coal industry. With the world's largest contributor to climate change in terms of accumulated emissions out, other countries might leave as well.

With a vacuum at the federal level in the US and other countries, cities and states are attempting to take mitigation into their own hands. C40 cities is a network of the world's megacities committed to addressing climate change, founded in 2005. Member cities must

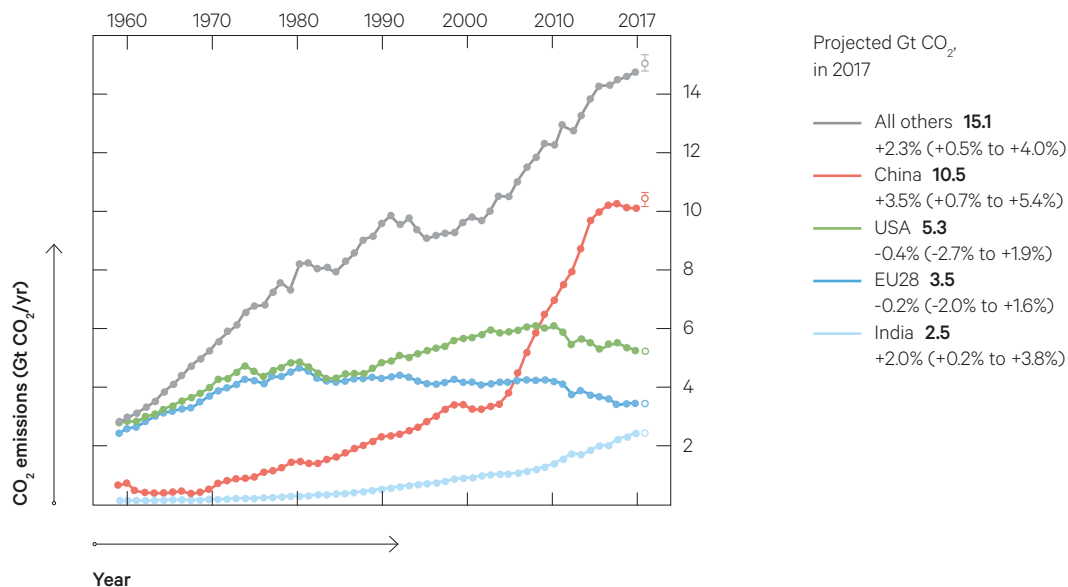


Figure 3. CO<sub>2</sub> emissions from fossil fuel and industry. Global emissions from fossil fuels and industry were projected to rise by 2% in 2017, stubbornly consistent with historical exponential emissions growth. Source: Global Carbon Project

have a plan to deliver their contribution to constraining warming to no more than 1.5°C. My home state of California, the world's fifth largest economy, is a leader within the US on climate action. In 2018, California passed groundbreaking legislation to get to 60% renewable electricity by 2030 and 100% carbon-free electricity by 2045.

**In 2017, the four largest emitters were China (28%), the US (15%), the EU (10%), and India (7%). China's emissions began increasing dramatically around 2000 and surpassed the US in 2005. However, China's emissions decreased by 0.3% from 2015 to 2017**

Renewables have been growing globally, driven largely by cost decreases and volume from China. In 2016, solar photovoltaic capacity grew by 50%; and in 2017 solar added 98 gigawatts globally, more than fossil fuel and nuclear combined. China now has more than 3 million jobs in solar, over 1,000 times the number of solar jobs in the US (230,000).<sup>24</sup> While the US had a chance to take the lead in the clean energy economy five or ten years ago, today China has a commanding lead. In 2017, just under 5% of the world's electricity was generated by solar and wind.<sup>25</sup>

China is also taking the lead in electric vehicles (EVs). Total EVs in China in 2017 numbered 3.1 million, up 56% from 2016.<sup>26</sup> In Norway in 2017, 39% of car sales were EVs. Ireland has pro-





claimed that it will not allow sales of internal combustion vehicles after 2030. In California, EVs accounted for almost 8% of light-vehicle sales in April of 2018.

While it is hard to measure cultural shift, I feel that the last decade has seen a significant cultural shift on climate. Grassroots movements and direct action are gaining momentum. The Keystone XL fight and Standing Rock were lines drawn in the sand against fossil fuel extraction; these and other actions are raising awareness that fossil fuel is harmful. In Europe, there is a major movement against airport expansion. Academics are attempting to shift their culture of frequently flying in airplanes. The media still usually fails to mention climate change after climate-related disasters, but this may be beginning to change. The climate litigation movement is beginning to gain traction, with youth plaintiffs suing the US government in what might turn out to be a watershed case. The Sierra club has stopped over two hundred new coal plants and secured 275 existing plants for retirement. My anecdotal sense is that people in the US are talking about climate change in their everyday lives more than they did in 2008, but still not enough.

The popular climate movement is transforming into the climate justice movement, a response to the deep injustice of climate breakdown: those who contributed to it the most (wealthy people in wealthy nations in the Global North) will suffer the least, and those who contributed the least (poor people in poor nations in the Global South) will suffer the most. Half of global emissions are produced by 10% of the global population. Climate justice could help move us to an inclusive “all of the above” mindset more quickly, creating the “movement of movements” that is needed. The climate justice movement must never lose sight of the most important metric: emissions. Greenhouse gas molecules do not care about our politics, and a powerful movement must include everyone. A principal focus on reducing emissions could help minimize the risk of further politicization of climate action.

## **In 2016, solar photovoltaic capacity grew by 50%; and in 2017 solar added 98 gigawatts globally, more than fossil fuel and nuclear combined**

In summary, the human response in all its inevitable complexity has begun to ramp up at every scale. It is currently insufficient, but a cultural shift is occurring. People are inspiring each other.

One recent night, a friend called to tell me that a nearby community’s city council would be voting in a couple of hours on whether to make their residents’ default electrical mix 50% renewable or 100% renewable. Here in California, many of our communities have been adopting “community choice energy” under a state law allowing cities and counties to band together and choose how their electricity mix is generated. Residents are enroled in a particular plan by default, and can opt into another plan. Since the vast majority remain in the default, our mission was to lobby the city council for the 100% renewable option. So I headed downtown. I spoke for two minutes, and the council began deliberations. My friend, running late from a similar action in another city, arrived just before the vote, and was given a chance to speak as well. The council voted 5–0 in favor of the 100% renewable default option. After the vote, three of the council members actually came down from the bench and exchanged hugs with us. It was a small action, but it felt wonderful.

Each one of us is a single mammal. We are limited in what we are able to do. It is therefore natural to feel overwhelmed by climate breakdown. But shutting down will only make us

feel worse. If we each do all we can, it may not be enough; but we cannot do more. Doing all we can connects us to everyone else around the world also doing all they can. It is the best cure for the climate blues.<sup>27</sup>



### The Carbon Budget for 1.5°C

Humanity as of 2017 is emitting  $42 \pm 3$  GtCO<sub>2</sub> per year, or about 1 GtCO<sub>2</sub> every 8.6 days,<sup>28</sup> up from 34 GtCO<sub>2</sub> per year in 2007 (the 2002–12 mean).<sup>29</sup> This implies humanity's CO<sub>2</sub> emissions have been increasing at about the same 2.2% annual rate as the atmospheric CO<sub>2</sub> fraction. Over these eleven years, the human population increased by 13%, or nearly a billion people (from 6.7 billion in 2007 to 7.6 billion in 2017). Global per capita CO<sub>2</sub> emissions increased by 9%. In other words, global emissions are driven by growth in both population and individual consumption.

By 2017, humanity had emitted about two trillion tonnes of CO<sub>2</sub> cumulatively ( $2200 \pm 320$  GtCO<sub>2</sub>).<sup>30</sup> According to the IPCC, a 1.5°C maximum is likely still physically possible. By the end of 2018, for a two-thirds chance of staying below 1.5°C the IPCC estimates that humanity can emit an additional 530 GtCO<sub>2</sub> or 380 GtCO<sub>2</sub>, depending on whether temperature is measured using sea surface temperatures or near-surface air temperatures over the ocean, respectively.<sup>31</sup>

At current rates of emissions, the first year the planet will surpass the 1.5°C mark would likely be in the late 2020s or early 2030s.<sup>32</sup> If emissions continued, the planet would surpass 2°C of warming around mid-century.

If humanity were to begin ramping down immediately and achieve net-zero CO<sub>2</sub> emissions by 2055, and in addition began ramping down non-CO<sub>2</sub> forcings by 2030, CO<sub>2</sub> emissions would remain within the 530 GtCO<sub>2</sub> budget.<sup>33</sup> This would require global cooperation and mobilization greater even than World War II, and no fighting between nations could be afforded.

**At current rates of emissions, the first year the planet will surpass the 1.5°C mark would likely be in the late 2020s or early 2030s. If emissions continued, the planet would surpass 2°C of warming around mid-century**

If humanity achieved net-zero CO<sub>2</sub> emissions and declining (or net-zero) non-CO<sub>2</sub> forcings on these approximate timescales, anthropogenic warming would likely halt after a few decades.<sup>34</sup> Earth system feedbacks, such as permafrost methane release, loss of forests, and the ice-albedo feedback, may continue to cause additional warming long after this point; the trajectory of that warming (its strength as a function of time, and its ultimate duration) is unknown.<sup>35</sup> A gradually decreasing fraction of accumulated atmospheric CO<sub>2</sub> would remain in the atmosphere for thousands of years, causing the global temperature to decline gradually as well; maintenance of anthropogenic warming on a timescale of centuries would cause additional ice loss and sea-level rise.<sup>36</sup> Carbon dioxide removal (CDR) could mitigate these long-term impacts, but the feasibility of large-scale CDR is unknown.

The relatively simple framing of CO<sub>2</sub> budgets hides much complexity and uncertainty, and the scientific community feels that the IPCC's estimate is more likely overly optimistic than overly pessimistic, for two principal reasons. First, the IPCC used a preindustrial baseline







The last decade has seen a revolution in *in situ* ocean measurements with the advent of Argo floats, a system of about 4,000 floats distributed over the global ocean

Antarctica, Scotia Sea, near South Georgia, waves crashing on tabular iceberg with cave





of 1850–1900. Because large-scale fossil CO<sub>2</sub> emissions began about a hundred years earlier, this late-nineteenth-century baseline could understate global mean temperatures by up to 0.2°C.<sup>37</sup> In the worst case, this would imply, then, that humanity was in arrears by about a decade, or roughly 400 GtCO<sub>2</sub>, leaving a budget of as little as 100 GtCO<sub>2</sub>. Second, the IPCC has not attempted to include carbon cycle feedbacks such as permafrost melting and wetland emissions; according to the IPCC SR1.5, emissions from these feedbacks could detract 100 GtCO<sub>2</sub> from the 1.5°C budget. Due to these two considerations alone, therefore, it is conceivable that humanity has already burned through the entire CO<sub>2</sub> budget for 1.5°C. Each passing day's emissions, of course, make "lock in" of 1.5°C more likely.

Additional sources of uncertainty in the budget include incomplete knowledge of the climate response ( $\pm 400$  GtCO<sub>2</sub>) and uncertainty about how humanity will mitigate non-CO<sub>2</sub> forcings (methane, black carbon, nitrous oxide, and hydrofluorocarbons,  $\pm 250$  GtCO<sub>2</sub>).

It is important to remember that 1.5°C is an arbitrary threshold for communicating risks and setting goals. As for what actually matters on the planet, the warmer it gets, the worse the impacts will be. Whatever the actual (essentially unknowable) amount of CO<sub>2</sub> we can emit while remaining below any arbitrary level of warming, be it 1.5°C or 2°C, humanity's course of action remains the same: reduce emissions as fast as possible (mitigation), and buckle in for a wild ride (adaptation).

### What We Stand to Lose: 1.5°C, 2°C, and Beyond

It does seem clear, however, that 1.5°C is the best humanity can possibly hope for at this point, and unfortunately this level of warming will intensify impacts beyond today's. Here I present a brief survey of impacts. Due to space constraints, I focus on comparing impacts at 1.5°C to impacts at 2°C; and although impacts are expected to exhibit deep regional variation, I focus here on the global picture.

Heat extremes over land are projected to shift to about 3°C hotter, although this shift will depend on region; generally speaking, expect a 47°C heatwave to clock in at about 50°C in the warmer world.<sup>38</sup> At 2°C, heatwaves will be 4°C hotter than they currently are; expect a 47°C heatwave to clock in at 51°C. Frequency of warm extremes (that would occur over land once in twenty years in today's climate) are projected to increase by 129% under 1.5°C of warming, and by 343% under 2°C of warming.<sup>39</sup> Warm spell durations are projected to increase on average by seventeen days under 1.5°C of warming, and by thirty-five days under 2°C of warming.<sup>40</sup>

The frequency of extreme rainfall over land is projected to increase by 17% under warming of 1.5°C and 36% under warming of 2°C.<sup>41</sup>

**Heat extremes over land are projected to shift to about 3°C hotter, although this shift will depend on region; generally speaking, expect a 47°C heatwave to clock in at about 50°C in the warmer world**

Sea level is projected to rise 26 cm to 77 cm by 2100 for 1.5°C of warming, and 2°C of warming is projected to bring an additional 10 cm; even this seemingly modest additional rise likely



translates into an additional ten million displaced people globally.<sup>42</sup> Annual economic losses due to foods are estimated at US\$10 trillion under 1.5°C of warming, and US\$12 trillion under 2°C of warming.<sup>43</sup> Sea level will continue to increase beyond 2100, with a potential for increases of several meters over hundreds of years as ice sheets are irreversibly lost.<sup>44</sup> However, the trajectory of ice-sheet loss and the resulting sea-level rise is still highly uncertain. Ice-sheet scientists are increasingly pointing out the potential for a much larger, much more rapid sea-level rise due to nonlinear loss of ice sheets, especially in west Antarctica, which could be triggered by global warming of even just 1.5°C.<sup>45</sup> Such a rapid increase would have profound implications for the world's coastal cities, the global economy, and global political stability.

Relative to 1.5°C, average vertebrate habitat loss doubles, and average invertebrate habitat loss triples at 2°C.<sup>46</sup> Relative to 1.5°C, an additional 1.5 million to 2.5 million square kilometers of permafrost are projected to melt at 2°C.<sup>47</sup>

At 1.5°C, between 70% and 90% of the world's warm water coral reefs are projected to be lost; at 2°C, 99% of reefs would be lost.<sup>48</sup> At 1.5°C, the global annual fish catch is projected to decrease by 1.5 million tonnes compared to 3 million tonnes for 2°C.<sup>49</sup>

The Atlantic meridional overturning circulation is projected to decrease by 11% under 1.5°C of warming and by 34% under 2°C of warming in 2100.<sup>50</sup>

Biodiversity losses are rapid and long-lasting; for example, biodiversity loss in mammals expected to occur over the next fifty years will last for millions of years, as evolution gradually recreates diversity.<sup>51</sup> I personally find it staggering that the consequences of the decisions we make over the next few decades will stretch out for millions of years.

Health risks increase with temperature. The suitability of drylands for malaria transmission is projected to increase by 19% under 1.5°C of warming and 27% under 2°C of warming.<sup>52</sup> As temperatures increase, there will be more numerous *Aedes* mosquitoes, and over a larger geographic range, increasing the risk of dengue fever, chikungunya, yellow fever, and Zika virus. The range and seasonality of Lyme and other tick-borne diseases are projected to expand in North America and Europe, and projections worsen with temperature.<sup>53</sup> At 1.5°C, twice as many megacities are likely to experience heat stress, exposing 350 additional people to deadly heat by 2050.<sup>54</sup>

## One study projects global maize yields to decrease by 6% under warming of 1.5°C and 9% under warming of 2°C by 2100

The integrated global food system is complex, to say the least. Future food security will depend on the interplay between regional crop stresses (temperature, water, pests, soil depletion), adaptation, population, consumption patterns, energy costs, and international markets. So far, technological yield gains have managed to keep pace with global warming, but this may not last as stresses increase. One study projects global maize yields to decrease by 6% under warming of 1.5°C and 9% under warming of 2°C by 2100.<sup>55</sup> Another study projects that mean yields in four countries responsible for over two thirds of maize production (the US, Brazil, Argentina, and China) will decrease by 8–18% under 2°C of warming, and by 19–46% under 4°C of warming.<sup>56</sup> (These estimates do not include additional stresses from aquifer depletion.)

It is possible that yields of all crops could drop more steeply under higher temperatures, as plant and pest responses to warming are not expected to remain linear; this steep drop





with temperature ensures that warming will bring greater harvest variability. With increasing likelihood of simultaneous yield losses in multiple regions and multiple crops, serious price shocks such as the tripling that occurred in 1972–1974 (due to extreme temperatures in the USSR) could become frequent.<sup>57</sup> When projected yield losses and variability increases are combined with projected population growth, it is a cause for concern.

Global poverty is projected to increase with warming. At 1.5°C, the number of people exposed to climate-related poverty is reduced by up to several hundred million by 2050 relative to 2°C.<sup>58</sup>

Beyond 2°C of warming, impacts get much worse and adaptation gets more expensive and less effective.

#### Near-Term Solutions: “All of the Above” but “Keep It Real”

Preventing more than 1.5°C of warming above the 1850–1900 baseline at this late date (if it is still possible) would require a massive global mobilization, far exceeding current levels of action and even current pledges of the Paris Agreement. In my opinion, to succeed, humanity will need to make climate action its highest priority, above even economic growth.

The recent IPCC special report on global warming of 1.5°C (IPCC SR1.5) suggests that such a goal is at least still physically possible.<sup>59</sup> Whether this is true or not, however, is immaterial to humanity’s course of action. The path to 2°C of warming and beyond is via 1.5°C of warming, and, as we saw in the previous section, impacts worsen profoundly as warming increases.

While the fine details of these projected impacts are still the subject of scientific debate, the essential fact of dangerous anthropogenic warming is not. The human response, however, will always be up for debate. The recommendations that follow are therefore necessarily my opinions as a global citizen.

**Preventing more than 1.5°C of warming above the 1850–1900 baseline at this late date (if it is still possible) would require a massive global mobilization, far exceeding current levels of action and even current pledges of the Paris Agreement. To succeed, humanity will need to make climate action its highest priority, above even economic growth**

Overall, we need to be very careful to adopt solutions that will result in actual emissions reductions, here and now. Call me crazy, but I think that a good way to respond to a crisis caused by burning fossil fuel is to stop burning fossil fuel, at every scale. We should not necessarily expect life to go on exactly as it has in the age of fossil fuel. And that’s OK: a livable planet is far more important than preserving a status quo that has not even been making us happy.

#### Dangerous Thinking

Assessments of the utility of a specific climate action must also include its feasibility. Unfortunately, human nature gravitates to potential solutions that require no immediate change,



but that appear somehow technologically sexy. Perhaps many of us tend to fetishize such techno-fixes because they resonate with one of our society's most powerful myths: progress. Such "solutions" are dangerous because they reduce urgency and divert from meaningful climate action.

Perhaps the most dangerous example of this sort of magical thinking is negative emissions technologies (NETs). Leading NET schemes include bioenergy carbon capture and storage (BECCS) and enhanced rock weathering. While I believe that research on NET schemes should be ramped up, counting on them to solve climate breakdown is dangerously irresponsible. These technologies do not exist at scale yet, and may not exist in time to help. To assume that NETs will someday draw down CO<sub>2</sub> from the atmosphere, thereby magically atoning for today's carbon, is to burden our children with both paying for the scheme (guaranteed by the second law of thermodynamics to be costly) *and* suffering with a higher level of warming than if mitigation had urgently proceeded.

## **Research on NET schemes should be ramped up, but counting on them to solve climate breakdown is dangerously irresponsible. These technologies do not exist at scale yet, and may not exist in time to help**

Another example is agricultural biofuel. Given the projections for food insecurity described in the previous section, there is no feasible path for agricultural biofuel to become a major component of climate action. Furthermore, the energy return on energy invested (EROEI) for corn ethanol is approximately 1:1, meaning it takes as much energy to produce corn ethanol as is released by burning corn ethanol.<sup>60</sup> Even inefficient tar sands oil has an EROEI of 4:1, and wind power has an EROEI of 20:1. While research should continue on unconventional energy sources such as artificial photosynthesis, these programs should not be seen as "solutions" for the simple reason that they do not yet exist, and may not exist in time to mitigate the current climate crisis. To mitigate the current climate crisis effectively, they would likely need to be deployed and scaled up to global levels within the next ten years or so (that is, before 2030); this seems exceedingly unlikely. The longer it takes for them to arrive, the less effective they will be.

Solar geoengineering refers to large-scale efforts by humanity to reflect a fraction of incoming sunlight.<sup>61</sup> Of the possible schemes, the most popular involves mimicking volcanic eruptions by distributing bright aerosols into the stratosphere. Unlike NETs, aerosol geoengineering is technologically and economically feasible. Unfortunately, it could cause additional disruptive climate and precipitation changes; it would not address ocean acidification; and worst of all, it could set up a "termination shock," saddling our children with a massive, sudden temperature spike if civilization becomes unable for any reason to continue supporting a large fleet of pollution-spraying airplanes. Our understanding of the ramifications of aerosol geoengineering is still crude, and research in this area should accelerate. It will be unfortunate if humanity feels that aerosol geoengineering is its best option, but this could soon be the case if procrastination continues and climate impacts are allowed to worsen. In this case it would be far better to have mature research in hand to guide the project.



Environmental activists gripping an iron fence at the White House during a protest against the Keystone XL pipeline in March 2014





Unlike the above pseudo-solutions, meaningful action will require cultural shift and broad public support. Burning fossil fuel causes harm, and must no longer be seen as acceptable. The public does not understand the urgency of the situation. To me, this is actually a source of hope: once the public does wake up, the rapidity of action could be genuinely surprising. Public support would cause powerful institutions and governments to fight for climate action, instead of against it. To me, the difference would be like night and day.

**Once the public wakes up, the rapidity of action could be genuinely surprising. Public support would cause powerful institutions and governments to fight for climate action, instead of against it. To me, the difference would be like night and day**

Hastening this cultural shift is up to each of us. There is no shortcut. I think this is an empowering message: personal actions matter, because individual, community, and collective action are inextricably intertwined. Collective action enables individual action (by changing systems) and individual action enables collective action (by changing social norms). The distinction between individual and collective action blurs, and it turns out to be a false dichotomy. We need action at all scales. For those who are concerned about climate breakdown, I recommend deeply and systematically reducing your own emissions, and working to spread that change within your community and beyond, via people you know, community institutions, and passionate communication aligning with your talents.<sup>62</sup>

In the very near term, the best first step any nation could take would be to implement a carbon fee and dividend.<sup>63</sup> Whenever any fossil fuel is taken from the ground or imported over a border, a fee would be assessed based on the embodied CO<sub>2</sub> emissions (and possibly also on other greenhouse gases). The fee would start at a modest level (likely less than \$100 per tonne CO<sub>2</sub>) and increase every year at a scheduled rate, eventually becoming prohibitively high (thousands of US dollars). This would fix a glaring problem in the market—that while climate pollution is incredibly costly to civilization, it is still free to pollute. Alternatives to fossil fuel would thus become increasingly viable throughout the economy, from carbon-free electricity, to locally grown organic food, to sustainable manufacturing, to slow travel options other than flying. The price predictability would accelerate a phased, orderly divestment from fossil fuel, and systematic investment in alternatives. People, institutions, and businesses would naturally also look for ways to use energy more efficiently (including food). Diets would naturally shift away from meat, and innovative meat substitutes would continue to be developed and embraced by consumers.

One hundred percent of the revenue would then be returned to every citizen. This would ensure that poor households would not be penalized by the policy; in fact, the poorest three quarters or so of households would come out ahead, since wealthy people use more energy, and would contribute proportionally more to the revenue. The dividend would make the policy popular, and the policy ought to be amenable to both conservatives (since it fixes the market, does not grow the government, and does not hurt the economy) and liberals (since it protects the environment and helps poor households and marginal communities). Imagine that: a climate action that unites.





Another excellent step any nation could take would be to simply use less energy. This would make the energy transition much easier. For example, I estimate that halving electricity usage in the US would require building only a quarter as much new carbon-free generative capacity; it would also support electrification of other sectors, such as transportation, industry, and heating and cooling. Using less energy would require government regulation, which, of course, requires public support to do in any meaningful way. The US state of California is an excellent example in this regard, and continues to demonstrate what is actually possible.

Dealing with non-CO<sub>2</sub> emissions (such as hydrofluorocarbons) and deforestation would also benefit from regulation. First, nations will need to decide to lead within their own borders. They will then need to advocate for strong, enforceable international regulations. In other words, the cultural shift will need to become strong enough to support meaningful action at the international level. To do this will require lead nations such as China, the US, and EU nations to recognize their outsized role in the climate crisis, and agree to provide appropriate economic support to countries who have had much smaller roles. For example, lead nations benefit if Indonesia stops deforestation; perhaps they should provide economic support for this, to offset Indonesia's financial losses. More generally, the cultural shift will need to become strong enough to begin atonement (recognition of responsibility) and reparations (economic support).

**Lead nations such as China and the US, as well as the EU, must recognize their outsized role in the climate crisis, and agree to provide appropriate economic support to countries who have had much smaller roles. Lead nations benefit if Indonesia stops deforestation; perhaps they should provide economic support for this, to offset Indonesia's financial losses**

In a democracy, major policy programs such as a carbon fee and dividend and powerful regulations to rapidly and systematically deal with non-CO<sub>2</sub> emissions require policy makers who support such programs; having those policy makers in office requires them winning elections on platforms that include such programs; winning those elections requires having sufficient votes; having sufficient votes requires a public that actively supports the policies; and having that supportive public requires a cultural shift. Polls in the US, for example, consistently show climate at or near the bottom of the list of issues of concern for both Republican and Democratic voters; this needs to change.

Therefore, the question “How do we get policies needed for rapid, meaningful action on climate change” is really the same as “How do we get a cultural shift to climate concern”?

Humanity also needs to move toward respecting ecological constraints as rapidly as possible. In my opinion, meaningful collective action that leads to civilizational respect for planetary boundaries will require the cultural shift to proceed even further than required for strong regulatory frameworks and carbon fees. In my personal vision for how humanity can avoid further civilizational collapse, I envision climate regulations and carbon fees happening first, and feeding back into an even larger cultural shift by inspiring hope and improving lives. The near-term, feasible step of a strong fee and dividend, for example, could catalyze a shift in narrative—from “joyless human dominion and growth” to “joyful stewardship on



The Golmud Solar Park in Qinghai, China, is one of that country's first six clean-energy projects. China invested over 127 billion dollars in renewable energy sources in 2017



a fragile and interconnected Earth.” It is difficult to imagine the transformative power such a shift could unleash.



### The Astronomical Perspective

When viewing the iconic photograph known as “Earthrise,” taken by Apollo 8 astronaut Bill Anders in 1968, one cannot help but see our planet as the fragile, astronomically unlikely oasis in a dark cold void that it actually is. We live on spaceship Earth, together with each other and the rest of life as we know it. The world that many of us and our collective culture take for granted, to “have dominion over,” is in reality a tiny, isolated system perfectly situated relative to a spherical nuclear inferno and with just the right mixture of chemicals in its paper-thin atmosphere to be congenial to human life and human civilization. Of all the accomplishments in climate science over the last ten years, perhaps this is the most important: adding to the unequivocal evidence of the fragility of Earth’s climate, and that humans are now singlehandedly moving it beyond the safe zone for civilization.

## Climate breakdown might help to explain the perplexing silence emanating from the billions and billions of planets orbiting almost every star in the night sky

Astrobiology is the study of how life arises on planets throughout the universe. Astrobiologists now speculate that climate breakdown may be a challenge not just for human civilization, but for civilizations on other planets, as well.<sup>64</sup> Climate breakdown might help to explain the perplexing silence emanating from the billions and billions of planets orbiting almost every star in the night sky. Perhaps ours is not the only civilization to discover fossil fuel and realize its usefulness, become addicted to it, and dangerously warm its planet’s climate.

It was natural that humanity made use of fossil fuel to build civilization. But now that we so clearly know that burning fossil fuel causes irreversible harm at a planetary scale, we must transition as though the future of human and nonhuman life on Earth depends on it.

### Toward a New Enlightenment: Humanity’s Place in the Web of Life

As humanity goes forward, it would do well to examine why it has arrived at this civilizational crossroads, and what lessons can be learned. The two most urgent crises facing the global biosphere, and therefore humanity, are climate change and habitat loss. Both of these arise from the prevailing mindset of the dominant globalized culture of “take and consume,” with little thought of how one’s actions might affect other beings on spaceship Earth, both human and nonhuman.

The main lesson to be learned, then, might be that everything in the biosphere—including any individual human, and certainly the human species—is not separate, but connected to everything else. This logically leads to a golden rule for sustainable civilization: treat the Earth and its beings as you would have it and them treat you—because they are you. A key difference between those with the mindset of separation and those with the mindset of connection is

that the former take for granted, and selfishness informs their actions; while the latter feel deep gratitude for all the factors of their existence, and this gratitude informs their actions.

In addition to a rapid transition away from fossil fuel, humanity must confront the other major ongoing global crises: habitat loss. At least half of the Earth must be designated for nonhuman habitat to make space for other species.<sup>65</sup> This might be the only way our planet can begin to regenerate a healthy biosphere, which is certainly in humanity's long-term self-interest. This would require an entirely new paradigm for humanity: a paradigm not of uncontrolled exponential growth, but, instead, a collective self-awareness, respect for limits, and a joyful humility at being just one strand of the web of life on this beautiful planet.





## Notes

1. A. Shepherd et al., "Mass balance of the Antarctic ice sheet from 1992 to 2017," *Nature* 558 (2018). <https://doi.org/10.1038/s41586-018-0179-y>. For a journalistic summary, see Kendra Pierre-Louis, "Antarctica is melting three times as fast as a decade ago," *New York Times*, June 13, 2018. <https://www.nytimes.com/2018/06/13/climate/antarctica-ice-melting-faster.html>.
2. C. D. Keeling, S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, and H. A. Meijer, "Exchanges of atmospheric CO<sub>2</sub> and <sup>13</sup>CO<sub>2</sub> with the terrestrial biosphere and oceans from 1978 to 2000. I. Global aspects," SIO Reference Series, No. 01-06, San Diego: Scripps Institution of Oceanography, 2001 (88 pages). [http://scrippsco2.ucsd.edu/data/atmospheric\\_CO2/primary\\_mlo\\_CO2\\_record](http://scrippsco2.ucsd.edu/data/atmospheric_CO2/primary_mlo_CO2_record).
3. P. Kalmus, *Being the Change: Live Well and Spark a Climate Revolution*, New Society Publishers, Gabriola Island, 2017, p. 42.
4. The data are from Berkeley Earth, Land+Ocean surface temperature time series. <http://berkeleyearth.org/data/>.
5. To calculate this more stable estimate for the 2017 temperature, I performed a linear fit from 2002 to 2017, extrapolated this fit to 2032, and took the mean of the resulting time series.
6. Data from NOAA, Ocean Climate Laboratory, Global Ocean Heat and Salt Content, "Basin time series of heat content (product, 0–2000 meters)," [http://hdc.noaa.gov/OC5/3M\\_HEAT\\_CONTENT/basin\\_data.html](http://hdc.noaa.gov/OC5/3M_HEAT_CONTENT/basin_data.html). Data described in S. Levitus et al., "World ocean heat content and thermocline sea level change (0–2000 m), 1955–2010," *Geophysical Research Letters* 39 (2012). <https://doi.org/10.1029/2012GL051106>.
7. Data: Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS). Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, "Uncertainty in modeled Arctic sea ice volume," *J. Geophys. Res.* 2011. <https://doi.org/10.1029/2011JC007084>.
8. W. Booth and A. Ferris-Rotman, "Russia's Suez Canal? Ships start plying a less-icy Arctic, thanks to climate change," *The Washington Post*, September 8, 2018. [https://www.washingtonpost.com/world/europe/russias-suez-canal-ships-start-plying-an-ice-free-arctic-thanks-to-climate-change/2018/09/08/59d50986-ac5a-11e8-9a7d-cd30504ff902\\_story.html](https://www.washingtonpost.com/world/europe/russias-suez-canal-ships-start-plying-an-ice-free-arctic-thanks-to-climate-change/2018/09/08/59d50986-ac5a-11e8-9a7d-cd30504ff902_story.html).
9. J. Watts, "Arctic warming: scientists alarmed by 'crazy' temperatures," *The Guardian*, February 27, 2018. <https://www.theguardian.com/environment/2018/feb/27/arctic-warming-scientists-alarmed-by-crazy-temperature-rises>.
10. P. R. Thompson et al., "Sea level variability and change" (in *State of the Climate in 2017*), *Bulletin of the American Meteorological Society*, 99(8), S84–S87 (2018).
11. C. M. Eakin et al., "Unprecedented three years of global coral bleaching 2014–17," (in *State of the Climate in 2017*), *Bulletin of the American Meteorological Society* 99(8), S74–S75 (2018).
12. T. Wallace et al., "Three of California's biggest fires ever are burning right now," *New York Times*, August 10, 2018. <https://www.nytimes.com/interactive/2018/08/10/us/california-fires.html>.
13. B. I. Cook et al., "Climate change and drought: From past to future," *Current Climate Change Reports* (2018). <https://doi.org/10.1007/s40641-018-0093-2>.
14. Y. Jiagan, "Response of subtropical stationary waves and hydrological extremes to climate warming in boreal summer," *Journal of Climate* (2018). <https://doi.org/10.1175/JCLI-D-17-04011>.
15. B. Hulac, "Climate change goes firmly in the 'loss' column for insurers," *Scientific American*, March 15, 2018. <https://www.scientificamerican.com/article/climate-change-goes-firmly-in-the-loss-column-for-insurers/>.
16. P. A. Stott, D. A. Stone, and M. R. Allen, "Human contribution to the European heatwave of 2003," *Nature* 432 (2004). <https://doi.org/10.1038/nature03089>.
17. Q. Schiermeier, "Droughts, heatwaves and floods: How to tell when climate change is to blame," *Nature* 560 (2018). <https://doi.org/10.1038/d41586-018-05849-9>.
18. V. Ramanathan and G. Carmichael, "Global and regional climate changes due to black carbon," *Nature Geoscience* 1 (2008).
19. Typically attributed to George Box.
20. J. Cook et al., "Consensus on consensus: A synthesis of consensus estimates on human-caused global warming," *Environmental Research Letters* 11 (2016). <https://doi.org/10.1088/1748-9326/11/4/048002>.
21. R. E. Benestad et al., "Learning from mistakes in climate research," *Theor Appl Climatol* (2016). <https://doi.org/10.1007/s00704-015-1597-5>. See also: Katherine Ellen Foley, "Those 3% of scientific papers that deny climate change? A review found them all flawed," *qz.com*, Sept. 5, 2017. <https://qz.com/1069298/the-3-of-scientific-papers-that-deny-climate-change-are-all-flawed/>.
22. CDIAC.
23. IPCC SR1.5.
24. E. Foehringer Merchant, "2017 was another record-busting year for renewable energy, but emissions still increased," *GreenTechMedia*, June 04, 2018. <https://www.greentechmedia.com/articles/read/2017-another-record-busting-year-for-global-renewable-energy-capacity#gs.lCE=AvM>.
25. World Nuclear Association, "World electricity production by source 2017," [www.world-nuclear.org/information-library/current-and-future-generation/nuclear-power-in-the-world-today.aspx](http://www.world-nuclear.org/information-library/current-and-future-generation/nuclear-power-in-the-world-today.aspx).
26. C. Gorey, "Global sales of EVs hit record number, but is it sustainable?" *Siliconrepublic*, May 30, 2018. <https://www.siliconrepublic.com/machines/global-sales-evs-2017>.
27. P. Kalmus, "The best medicine for my climate grief," *YES! Magazine*, August 9, 2018. <https://www.yesmagazine.org/mental-health/the-best-medicine-for-my-climate-grief-20180809>.
28. IPCC SR1.5.
29. IPCC AR5 WG1. To forestall potential confusion, note that 1 tonne of carbon emitted is equal to 3.67 tonnes of CO<sub>2</sub> emitted.
30. IPCC SR1.5.
31. Ibid.
32. Ibid.
33. Ibid.
34. Ibid.
35. Ibid.
36. Ibid.
37. A. P. Schurer et al., "Importance of the preindustrial baseline for likelihood of exceeding Paris goals," *Nature Climate Change* 7 (2017). <https://doi.org/10.1038/nclimate3345>.
38. IPCC SR1.5.
39. V. V. Kharin et al., "Risks from climate extremes change differently from 1.5°C to 2.0°C depending on rarity," *Earth's Future*. <https://doi.org/10.1002/2018EF000813>.
40. T. Aeronson et al., "Changes in a suite of indicators of extreme temperature and precipitation under 1.5 and 2 degrees warming," *Environmental Research Letters* 8 (2018). <https://doi.org/10.1088/1748-9326/aaaf6d>.
41. Ibid.
42. IPCC SR1.5.
43. S. Jevrejeva et al., "Flood damage costs under the sea level rise with warming of 1.5°C and 2°C," *Environmental Research Letters* 13 (2018). <https://doi.org/10.1088/1748-9326/aacc76>.
44. IPCC SR1.5.
45. Ibid.
46. Ibid.
47. Ibid.
48. Ibid.
49. Ibid.
50. J. B. Palter et al., "Climate, ocean circulation, and sea level changes under stabilization and overshoot pathways to 1.5 K warming," *Earth System Dynamics* 9 (2018). <https://doi.org/10.5194/esd-9-817-2018>.
51. Matt Davis, Søren Faurby, and Jens-Christian Svenning, "Mammal diversity will take millions of years to recover from the current biodiversity crisis," *Proceedings of the National Academy of Sciences* 2018. <https://doi.org/10.1073/pnas.1804906115>.
52. J. Huang et al., "Drylands face potential threat under 2C global warming target," *Nature Climate Change* 7 (2017). <https://doi.org/10.1038/nclimate3275>.

53. IPCC SR1.5.
54. Ibid.
55. C. Tebaldi and D. Lobell, "Differences, or lack thereof, in wheat and maize yields under three low-warming scenarios," *Environmental Research Letters* 13 (2018). <https://doi.org/10.1088/1748-9326/aaba48>.
56. M. Tigchelaar, et al., "Future warming increases probability of globally synchronized maize production shocks," *Proceedings of the National Academy of Sciences* 115 (2018). <https://doi.org/10.1073/pnas.1718031115>.
57. Ibid.
58. IPCC SR1.5.
59. Ibid.
60. Peter Kalmus, *Being the Change: Live Well and Spark a Climate Revolution*, New Society Publishers, Gabriola Island, BC, 2017, p. 98.
61. D. Dunne, "Explainer: Six ideas to limit global warming with solar geoengineering," Carbon Brief, September 5, 2018. <https://www.carbonbrief.org/explainer-six-ideas-to-limit-global-warming-with-solar-geoengineering>.
62. For more on why to change, how to change, and how to spread that change, see my book, *Being the Change: Live Well and Spark a Climate Revolution*.
63. To learn more, I recommend exploring resources at Citizens' Climate Lobby. <https://citizensclimatelobby.org>.
64. Adam Frank, *Light of the Stars: Alien Worlds and the Fate of the Earth*, W. W. Norton & Company, New York, 2017.
65. E. O. Wilson, *Half Earth: Our Planet's Fight for Life*, Liveright, New York, 2017.



**Ernesto Zedillo Ponce de León**  
Yale University

Ernesto Zedillo is the Director of the Yale Center for the Study of Globalization; Professor in the Field of International Economics and Politics; Professor of International and Area Studies; and Professor Adjunct of Forestry and Environmental Studies at Yale University. After almost a decade with the Central Bank of Mexico he served as Undersecretary of the Budget, Secretary of Economic Programming and the Budget, and Secretary of Education, before serving as President of Mexico from 1994–2000. He is Chairman of the Board of the Natural Resource Governance Institute and the Rockefeller Foundation Economic Council on Planetary Health and Co-Chair of the Inter-American Dialogue and is a member of The Elders. He serves on the Global Commission on Drug Policy and the Selection Committee of the Aurora Prize for Awakening Humanity and from 2008 to 2010 he was Chair of the High-Level Commission on Modernization of World Bank Group Governance. He is a Member of the Group of 30. He earned his BA degree from the School of Economics of the National Polytechnic Institute in Mexico and his MA and PhD at Yale University.

Recommended Book: *Why Globalization Works*, Martin Wolf, Yale University Press, 2004.

**For globalization to deliver to its full potential, all governments should take more seriously the essential insight provided by economics that open markets need to be accompanied by policies that make their impact less disruptive and more beneficially inclusive for the population at large. The real dilemmas must be acknowledged and acted upon, and not evaded as is done when tweaking trade policy is wrongly alleged to be the instrument to address unacceptable social ills.**



This article was written ten years after the outbreak of the worst crisis that the global economy had known in more than seventy-five years. The meltdown of the subprime market that had happened in the summer of 2007 in the United States became a full-blown crisis as Lehman Brothers collapsed in the early hours of the morning on September 15, 2008. The financial panic lived during those days not only marked the end of the so-called Great Moderation, but also the beginning of a period if not of twilight, at least of seriously deflated expectations, about modern globalization.

The decade that preceded the great crisis of 2008–09 was by several measures a golden period for globalization, which had been painstakingly rebuilt over the previous fifty years after its destruction during the Great Depression and World War II. Despite the Asian crisis of 1997–98, and other financial crises in other emerging economies, globalization intensified markedly in the 1990s to the point that already by the end of the twentieth century it had surpassed, at least on the trade and financial fronts, that phenomenon's previous golden era of a century earlier.

During the mini golden era of contemporary globalization, it was not only that trade in goods and services as well as capital flows across borders grew to unprecedented levels, but also that a process of economic convergence between the developed and the emerging and developing economies took place at last.

For over one hundred years, the group of countries known in recent history as the advanced ones—chief among them the United States, and those in western Europe and Japan—consistently generated sixty percent or more of global output. This group's large share of world production seemed to be perpetual. That economic predominance was not challenged even by the industrialization of the Soviet Union nor by the take off in the 1960s of some previously underdeveloped countries.

In 1950 the share of the advanced countries was sixty-two percent of global GDP, in purchasing power parity (PPP) terms, and twenty-two percent of world population. Two decades later that share of world output was the same and was still similar by 1990, notwithstanding that those countries' population had fallen to fifteen percent of the world's total (Maddison, 2001). Indeed, economic convergence of developing countries with industrialized ones appeared unachievable throughout most of the twentieth century. Countries accounting for most of the world's population seemed to be perennially condemned to only a small share of global GDP.

## **The decade that preceded the great crisis of 2008–09 was by several measures a golden period for globalization, which had been painstakingly rebuilt over the previous fifty years after its destruction during the Great Depression and World War II**

That seemingly historical regularity ended during the precrisis decade. Now, since the middle of the first decade of this century, the group of emerging or developing countries produces more than half of world output (Buiter and Rahbari, 2011). Needless to say, the richest countries' per capita income still surpasses every one of today's fastest growing emerging countries by a substantial margin. But the historical gap has closed significantly.

Part of this story of economic convergence is that during the last few decades the group of rich countries registered slower growth than before. However, convergence has been much more about the faster growth of developing countries, and this growth has been driven by





precisely those countries that, having been relatively closed a few decades ago, took the crucial step around the 1980s to integrate into the global economy. Thus, in less than a quarter of a century, a group of developing countries—home to more than fifty-five percent of the world's population—while doubling their ratio of trade to GDP and becoming more open to Foreign Direct Investment (FDI), were able to raise their per capita GDP at more than twice the rate of rich countries. Significantly, they also reduced both the number and proportion of their people living in extreme poverty—despite their substantial population increases.

Those countries are typically the ones that have been able to fast-track their industrialization by inserting their productive capacities into the global supply chains made possible by the information technology (IT) revolution (Baldwin, 2014).

Prior to this revolution, industrialization was about economies of scale as well as vertical integration and clustering of production processes. Consequently, building a competitive industry required a deep industrial base, a condition historically achieved by a rather small number of countries.

In turn, international trade was about specialization in the production of goods or commodities and essentially consisted of selling the merchandise produced in one country to customers in another; in other words, practically a two-way trade.

As computing and telecommunication capabilities became cheaper and enormously potent, in the presence of already inexpensive transportation costs and lower impediments to cross-border trade, it became economically attractive to separate the previously integrated and concentrated production processes. Production dispersion in internationalized supply chains now became cost effective and eventually, in many cases, the only way to remain competitive.

Increasingly, the old manufacturing clusters have given way to geographical fragmentation of production supported by incessant flows of investment, technology, personnel expertise, information, finance, and highly efficient transportation and logistics services, none of which would be attainable at the speed and with the certitude required without modern IT. This revolution has made it relatively inexpensive to coordinate complex activities situated in locations all over the world, making international supply chains feasible and profitable.

The implications of this transformation for the international division of labor are far reaching. On the one hand, by off-shoring fragments of their productive activities, the developed countries' firms can now put their more advanced technologies together with the low-cost labor of developing countries to augment their competitiveness. On the other, developing countries, by virtue of assimilating off-shored links of the supply chain, can now industrialize more rapidly without waiting to build the deep industrial base formerly required. Thanks to this unbundling and off-shoring, nations can industrialize, not by building, but by joining a supply chain, making industrialization faster and easier.

The new organization of production, driven by the Internet and the other tools of IT, does not pertain only to large corporations as commonly believed. The Internet is fueling transformations covering the entire value chain in practically all sectors and types of companies. In fact, its impact has been most significant in small- and medium-sized enterprises and start-ups. Remarkably, it is now possible for a small firm to be a global company practically as soon as it is born.

On the international trade front, increasingly, countries' comparative advantage is no longer about finished goods or commodities; it is about the finer tasks that make up the manufacturing, commercial, and financial processes necessary to ultimately produce and deliver the goods demanded by consumers. Interestingly, the services or tasks that go before and after the fabrication itself of each final good have become a larger proportion of its definitive value—this determines the so-called smile curve.



Three workers walking inside Piaggio Vietnam in April 2015. This factory on the outskirts of Hanoi produces the iconic Vespa, and has made over half-a-million scooters since the company moved its Asian headquarters from Singapore to Vietnam in 2009





Each good sold at the end of its supply chain is a conjunction of many countries' capital, labor, technology, infrastructure, finance, and business environments. This is leading to a profound change in the way we look at, study, and measure the evolution of the global economy.

Of course, the fact that technological progress is leading to a fundamental change in the pattern of production and trade—and with it a redistribution of economic might—across the world is not unprecedented in human history. It happened before with the Industrial Revolution. A profound shift and concentration of economic power among the nations of the world took place over the course of just a few decades, and those countries that played the new game best became the advanced ones, not only of the late nineteenth century but also of the twentieth.

In the economic rebalancing occurring during our time, although there have been many developing countries achieving rates of economic growth above those of rich countries, the case of China stands out among all of them. Thanks to its high average GDP growth for over two decades, ten years or so ago China had already become the second largest economy in the world, whereas as recently as 1990, it was only the tenth largest with a GDP even smaller than that of Spain that year.

By the eve of the financial crisis, it had also passed from being a marginal participant in global trade flows to be the largest exporter and the second largest importer of goods in the world, as well as the fastest growing importer of commercial services, ranking the third largest in the world. It also became the recipient of the largest flows of FDI, even surpassing the net flows going into the United States.

## **The services or tasks that go before and after the fabrication itself of each final good have become a larger proportion of its definitive value—this determines the so-called smile curve**

China's growth has been an accelerator of the new pattern of international production and trade that created unprecedented opportunities for other developing countries while allowing developed ones to have new fast-growing outlets for their own products, investments, and technologies. That growth also enlarged the pool of global savings, thus helping to loosen financial constraints, not least for the United States. Ironically, the latter aspect of China's success was also part of the story that led to the financial crisis that interrupted the mini golden era of globalization. It is now commonly believed that the crisis was caused by recklessness alone on the part of private financial institutions, mainly US but also European ones, and it seems to be forgotten that, in truth, the turmoil had some deep macroeconomic policy mismanagements and imbalances as its primary causes.

Lax fiscal and monetary policies played into the US economy's seemingly insatiable absorption of the vast pool of foreign savings, which, in turn, was made possible by key rigidities in the other major countries' economic policies, certainly China's, but others such as Germany and Japan as well. The underpricing of highly risky assets was the end result not only of faulty financial engineering but, more fundamentally, of too much liquidity chasing too few sound investment opportunities. Quite aside from the well-documented incompetent and foolhardy behavior of a number of financial institutions, without the massive borrowing by some countries and the massive lending by others, and of course the policies and structural factors underlying such imbalances, it would have been impossible to create such a tremendous economic disaster.



As warned repeatedly by some observers, the global macroeconomic imbalances were bound to cause trouble, and they did. Although the crisis originated and spread from the US financial markets, it soon became apparent that no significant economy would be spared the pain, and actually the guilt, from having allowed the roots of the crisis to grow so strong. For a while, the members of the Eurozone proclaimed themselves as victims and not culprits for the disaster on the basis that they had managed to keep a nearly balanced current account for the Union as a whole. They were failing to acknowledge that serious macroeconomic imbalances did exist within the European Monetary Union (EMU)—those between its northern members, chiefly Germany, and their southern partners. The truth is that Germany's current account surpluses were, among other things, feeding consumption binges in Greece, supporting exuberant construction booms in Spain and Ireland, funding unsustainable fiscal deficits in Portugal, and even helping to inflate the real-estate bubble in the US—as more than a few of the German banks' balance sheets painfully revealed in due time. Japan was another country that failed to take into account the effect of its large surpluses on its trading partners.

Extravagant claims about being decoupled from the US travails were also foolishly entertained in some important countries of Latin America. The commodities super-cycle that more or less survived until 2014 was the opioid that caused the leaders of those countries to rest on their laurels and fail to recognize the illnesses that had infected our economies well before the crisis. The chief consequence of the Latin American complacency of a few years ago is that, as the global economy gained enough momentum to leave the great crisis behind, the opposite happened in our region.

Sensibly, if only after the fact, the G20 leaders were right on target when, at their first Washington Summit of November 15, 2008, they identified insufficiently coordinated macroeconomic policies at the root of the crisis that had erupted with great force that fall. They recognized that as their national economies had become more interdependent, which had been positive for growth, this interdependence had also exacerbated policy challenges, including the need for more, not less, macroeconomic policy coordination. Unfortunately, that admission and the pledge to fix it, were made too late and were short lived.

The world has not and will not be the same after the other Black Monday, the one of September 15, 2008. For one thing, not only did the great crisis cause a meaningful loss of output throughout the years of its acute phase, it also brought about a negative effect on the trajectory of world output that has proved permanent. A secular dampening on global growth is part of our new normal. We are living through a period—one that will probably last a long time—of deflated or diminished expectations.

It is evident that the prospects for most economies, even in the presence of the relatively benign world output growth figures of 2017 and 2018, are very different from the ones entertained only a bit longer than a decade or so ago.

Although the list of factors suspected of contributing to the deflation of global growth expectations is not an insignificant one, not least as it includes both the mystery of reduced growth productivity as well as the aging of the labor force in advanced countries, particular consideration must be given to the question of whether globalization—a significant growth engine—might have peaked already and could even be at risk of significant reversion.

Naturally, most of the attention to the question of possible deglobalization has centered on trade (Hoekman, 2015; and IMF, 2016). The global trade to GDP ratio grew from roughly twenty-five percent in 1960 to sixty percent in 2008. This happened because, from 1960 to





the eve of the crisis in 2007, global trade in goods and services grew at an average real rate of about six percent a year, which was about twice that of real GDP growth during the same period. After a sharp drop during the crisis and a brief rebound in its immediate aftermath, trade growth has been very weak relative to the past; in fact, until 2017 it was not even keeping up with global output growth over several years. If that trend were to prevail, then the trade/GDP ratio of sixty percent would prove to be a peak and would give credence to the presumption that globalization is stalling and even risks reversing.

The confirmation of this presumption should be hugely concerning for those, like myself, who believe that the payoff of trade expansion has been on balance quite favorable not only for global growth—of both developed and emerging economies—but also in particular to increase average per capita income, reduce poverty rates, and accelerate human development in many developing countries. We get some relief regarding this issue from those who submit and empirically support the view that for the most part the trade slowdown has been driven essentially by cyclical factors, such as the weakness in aggregate demand caused in turn by the necessary rebuilding of balance sheets, which certainly has been the case for several years in the Eurozone and more recently even in China and other emerging economies.

**Although the list of factors suspected of contributing to the deflation of global growth expectations is not an insignificant one, particular consideration must be given to the question of whether globalization—an important growth engine—might have peaked already and could even be at risk of significant reversion**

Moreover, there are questions as to whether the process of global integration may also be stalling by virtue of the process of financial deglobalization that has occurred over the last ten years as gross cross-border capital flows decreased sixty-five percent (MGI, 2017). As in the case of trade, we are told that there is no cause for alarm since most of the contraction of international lending can be accounted for by the global retrenchment of European and a few US banks, which, to respond to credit losses, had to cut lending and other assets abroad enormously. From this perspective, the observed financial deglobalization, far from being a broad phenomenon, would reflect for the most part a cyclical deleveraging, by itself a necessary and actually benign evolution.

Be that as it may, even analyses more supportive of the cyclical nature of the trade slowdown acknowledge that there might be other factors at play that should not by any means be overlooked. That noncyclical, structural factors help to explain the trade slowdown is suggested by the fact that it actually started before the crisis—around the mid-2000s.

Among those factors there are some that should not be worrisome as they reflect evolutions that should have been expected, such as the completion of the phase of the fast integration of China and the central and eastern European economies into the global economy, a transition that by definition could not go on forever. Another would be that the international fragmentation of production fostered by the development of global supply chains has reached a plateau consistent with the existent IT and transportation technologies, a circumstance that may change as these technologies continue to make sufficient progress in the years to come.



**The commitments assumed by G20 leaders at their first summit in Washington came too late and were too short-lived**

Then president of the United States, George W. Bush, welcomes chancellor Angela Merkel to the White House before a dinner for G20 participants in November 2008





**Thanks to its high average GDP growth for over two decades, ten years or so ago China had already become the second largest economy in the world, whereas as recently as 1990, it was only the tenth largest**

An online-sales start-up employee in Beijing's Soho Galaxy takes a break during the night shift



But there are other noncyclical circumstances that should be of true concern. One is, of course, that the multilateral efforts to further liberalize trade have failed terribly for many years, not least with the Doha Round, now totally defunct despite the multiple pledges to complete it made by the G20 in the aftermath of the crisis. Another is the increase in protectionism that rather quietly—in a murky way, avoiding large-scale increases in the average level of border protection—took place over several years, again despite the solemn pledges of the G20 (Global Trade Alert).

The failure of further multilateral liberalization and the occurrence of creeping protectionism were bad enough for the prospects of global growth, but a much worse scenario has now emerged as a consequence of the trade wars that are apparently being actively pursued by the government of none other than the major economic power of the world, the United States. This is a scenario that, unthinkable until recently, now seems to be materializing.

The election and the actions of an old-fashioned, nationalistic, and populist government in the United States, the country that has championed and benefited the most from globalization, is the most significant downside risk faced by the world economy, a risk that has been grossly overlooked by financial markets at least until the fall of 2018.

It is not only the trade and investment consequences of the US neo-mercantilism that should raise serious concerns about the future of globalization and global growth. Equally or even more concerning is the country's use of nationalistic and populist postures at the expense of multilateral diplomacy in dealing with serious geopolitical issues, an approach that conceivably could make bellicose situations more likely with dire consequences for the world economy.

The crisis and its economic and political sequels have exacerbated a problem for globalization that has existed throughout: to blame it for any number of things that have gone wrong in the world and to dismiss the benefits that it has helped to bring about. The backlash against contemporary globalization seems to be approaching an all-time high in many places including, the United States.

Part of the backlash may be attributable to the simple fact that world GDP growth and nominal wage growth—even accounting for the healthier rates of 2017 and 2018—are still below what they were in most advanced and emerging market countries in the five years prior to the 2008–09 crisis. It is also nurtured by the increase in income inequality and the so-called middle-class squeeze in the rich countries, along with the anxiety caused by automation, which is bound to affect the structure of their labor markets.

Since the Stolper-Samuelson formulation of the Heckscher-Ohlin theory, the alteration of factor prices and therefore income distribution as a consequence of international trade and of labor and capital mobility has been an indispensable qualification acknowledged even by the most recalcitrant proponents of open markets. Recommendations of trade liberalization must always be accompanied by other policy prescriptions if the distributional effects of open markets deemed undesirable are to be mitigated or even fully compensated. This is the usual posture in the economics profession. Curiously, however, those members of the profession who happen to be skeptics or even outright opponents of free trade, and in general of globalization, persistently “rediscover” Stolper-Samuelson and its variants as if this body of knowledge had never been part of the toolkit provided by economics.

It has not helped that sometimes, obviously unwarrantedly, trade is proposed as an all-powerful instrument for growth and development irrespective of other conditions in the economy and politics of countries. Indeed, global trade can promote, and actually has greatly fostered, global growth. But global trade cannot promote growth for all in the absence of other policies.





The simultaneous exaggeration of the consequences of free trade and the understatement—or even total absence of consideration—of the critical importance of other policies that need to be in place to prevent abominable economic and social outcomes, constitute a double-edged sword. It has been an expedient used by politicians to pursue the opening of markets when this has fit their convenience or even their convictions. But it reverts, sometimes dramatically, against the case for open markets when those abominable outcomes—caused or not by globalization—become intolerable for societies. When this happens, strong supporters of free trade, conducted in a rules-based system, are charged unduly with the burden of proof about the advantages of open trade in the face of economic and social outcomes that all of us profoundly dislike, such as worsening income distribution, wage stagnation, and the marginalization of significant sectors of the populations from the benefits of globalization, all of which has certainly happened in some parts of the world, although not necessarily as a consequence of trade liberalization.

Open markets, sold in good times as a silver bullet of prosperity, become the culprit of all ills when things go sour economically and politically. Politicians of all persuasions hurry to point fingers toward external forces, first and foremost to open trade, to explain the causes of adversity, rather than engaging in contrition about the domestic policy mistakes or omissions underlying those unwanted ills. Blaming the various dimensions of globalization—trade, finance, and migration—for phenomena such as insufficient GDP growth, stagnant wages, inequality, and unemployment always seems to be preferable for governments, rather than admitting their failure to deliver on their own responsibilities.

## **Governments prefer to blame different aspects of globalization—trade, finances, and immigration—for phenomena such as insufficient GDP growth, stagnant wages, inequality, and unemployment rather than admitting their failure to deliver on their own responsibilities**

Unfortunately, even otherwise reasonable political leaders sometimes fall into the temptation of playing with the double-edged sword, a trick that may pay off politically short term but also risks having disastrous consequences. Overselling trade and understating other challenges that convey tough political choices is not only deceitful to citizens but also politically risky as it is a posture that can easily backfire against those using it.

The most extreme cases of such a deflection of responsibility are found among populist politicians. More than any other kind, the populist politician has a marked tendency to blame others for his or her country's problems and failings. Foreigners, who invest in, export to, or migrate to their country, are the populist's favorite targets to explain almost every domestic problem. That is why restrictions, including draconian ones, on trade, investment, and migration are an essential part of the populist's policy arsenal. The populist praises isolationism and avoids international engagement. The "full package" of populism frequently includes anti-market economics, xenophobic and autarkic nationalism, contempt for multilateral rules and institutions, and authoritarian politics.

Admittedly, only exceptionally, individual cases of populist experiments may become a serious threat to the process of global interdependence. When countries have toyed, democratically or not, with populist leadership, the damage has been largely self-inflicted, with any spillover effects limited to their immediate neighbors.



For example, Latin America is a place where populism has been pervasive at times. Yet, most of the hardship populism caused has been contained within the countries suffering the populist maladies. Unfortunately, a major exception to the rule of contained spillovers may be the current case of the United States, where the negative consequences of its leadership's neo-mercantilist stance could be enormously consequential for globalization, economic growth—including its own—and international peace and security.

As this paper was being written, the current US government has provided ample evidence, rather aggressively, of its protectionist and anti-globalization instincts. There was, of course, the very early decision by the Trump administration to withdraw from the Trans-Pacific Partnership (TPP), an action never really satisfactorily justified by the US president or any member of his cabinet. The decision proved rather ironic given that the TPP was an agreement molded to a great extent to favor American interests, not only on trade but also on matters such as intellectual property rights, investor-state arbitration, and labor standards.

There was also the action to initiate the renegotiation of the North American Free Trade Agreement (NAFTA) on false—or at best wrongheaded—premises. In May 2017, when the formal announcement to start the renegotiation process was made, the United States Trade Representative (USTR) argued that the quarter-century-old agreement no longer reflected the standards warranted by changes in the economy. This may have sounded plausible before noticing that the to-do list to update the agreement had already been addressed in the discarded TPP, of which both Mexico and Canada were a part. If NAFTA had been modernized in practice through the TPP, why call for renegotiation of the former while trashing the latter?

## **Latin America is a place where populism has been pervasive at times. Yet, most of the hardship populism caused has been contained within the countries suffering the populist maladies. Unfortunately, a major exception may be the current case of the United States**

The US government's duplicitous approach in dealing with its allies and trade partners was confirmed when the USTR published—as required by law—the objectives for the renegotiation (USTR, 2017). That document falsely associated NAFTA with the explosion of US trade deficits, the closure of thousands of factories, and the abandonment of millions of American workers. Frankly, the Mexican and Canadian governments should not even have sat down at the negotiating table without first receiving some apologetic explanation from their US counterparts about those unwarranted arguments. Accepting to negotiate on deceptive premises might help to explain why so little progress had been made after almost one year of talks.

Betting in mid-July of 2018 for a conclusion of the renegotiation of NAFTA within the targeted timeframe would have looked like an overwhelmingly losing proposition. After seven rounds of negotiation, the last one having taken place as far back as February 2018 with little or no progress, and then followed by several months of deadlock and even rhetorical confrontation, things started to change positively as August approached.

The deadlock was quite understandable. The US trade representatives had not moved a single inch from their most outlandish demands, giving credence to the idea that what they were seeking was to get a deal that, far from promoting, would have destroyed trade and investment among the NAFTA partners. Fortunately, the Canadian and Mexican governments did not cave to the US government's pretension. Repeatedly those countries' chief negotiators



A mural supporting Hugo Chávez in Caracas. The photo was taken during Venezuela's local elections in November 2008, ten years after that leader of 21st-century socialism came to power





expressed firmly and credibly that they would rather take the unilateral termination of NAFTA by the United States than sign an agreement that would have the same practical consequence.

It is not known what motivated the US government to move away from most of the recalcitrant positions it had held for almost a year (Zedillo, 2018). The important fact is that it did, leading to a deal first with Mexico on August 27 and then with Canada in the last hours of September 30, 2018.

There was the US insistence on a sunset clause that would automatically end the new trade agreement every five years unless the three governments agreed otherwise, a feature that would have precluded the certainty for investors that these deals are supposed to provide. They settled for a rather convoluted formula that avoids the sudden death of the agreement and makes possible—and practically certain—an extended life for it.

The US negotiators had demanded to make the NAFTA Investor State Dispute settlement procedure optional for the United States, with a view to deny such protection to its own companies, thus discouraging them from investing in the NAFTA partners. This demand was rejected all along by Mexico on the correct basis that it is important to give foreign investors every assurance that they would not be subject to discriminatory or arbitrary actions if they decided to invest in the country.

## **The Trump administration's duplicitous approach in dealing with its allies and trade partners was confirmed when the USTR published the objectives for the renegotiation (USTR, 2017). That document falsely associated NAFTA with the explosion of US trade deficits, the closure of thousands of factories, and the abandonment of millions of American workers**

The USTR was never shy about its dislike for the NAFTA investment rules, sometimes even questioning why it was a good policy of the United States government to encourage investment in Mexico. There are, of course, many good answers to this question, not least that by investing in Mexico, US firms, in order to do some part of their fabrication processes at a lower cost, get to be more competitive not only in the entire region but also globally, allowing them to preserve and enhance job opportunities for their American workers. Consequently, it is good for the two countries that the mechanism to protect American investments in Mexico was preserved despite the US negotiators' originally declared intentions.

By the same token, the US had sought to eliminate the dispute resolution procedure which protects exporters against the unfair application of domestic laws on anti-dumping and countervailing duties. This was a deal breaker for Canada, where there is the justified sentiment that the US has in the past abused the application of such measures against Canadian exporters. Canada's perseverance paid off and its exporters will have recourse to the dispute settlement system as it is in NAFTA.

The US side had also been stubborn about getting the Mexican side to accept in the new deal a special mechanism by which the US could easily apply anti-dumping tariffs on Mexican exports of seasonal fruits and vegetables. Mexico would not assent to the inclusion of this mechanism, and in the end the new agreement will not contain it—to the benefit of both American consumers and Mexican producers. Similarly, it is to the benefit of Canadian





consumers and US exporters of dairy products that Canada ultimately accepted an American request for at least a modest opening of such a market.

The only significant US demand accommodated by Mexico and Canada was in the automotive sector where more restrictive and cumbersome rules of origin are to be adopted. It has been agreed that seventy-five percent of a car or truck should have components from North America to qualify for tariff-free imports, up from the current level of 62.5 percent. Furthermore, seventy percent of the steel and aluminum used in that sector must be produced in North America, and forty percent of a car or truck would have to be made by workers earning at least \$16 per hour, a measure obviously calculated to put a dent in Mexico's comparative advantage. Fortunately, the destructive effects of the new rules of origin for trade and investment could be mitigated, in the case of cars, by the provision that vehicles failing to fulfill those rules would simply pay the low most-favored-nation tariff of 2.5 percent as long as total exports do not exceed an agreed reasonable number of vehicles.

Other things being equal, however, it is clear that the new regime will reduce both the regional and global competitiveness of the North American automotive industry, a result that will not be good for American, Canadian, or Mexican workers. Of course, other things may not be equal if the US government decides to impose tariffs, as it has threatened to do, on vehicles produced by European or Asian companies. If the US government were to impose those tariffs, the burden of the new regime would fall disproportionately on the American consumer.

As purported from day one, the trade agreement will be subject to an update on a number of topics such as digital trade, intellectual property rights, environmental policies, and labor practices. Interestingly the agreed new provisions really are a "cut-and-paste" of what was contained in the TPP, that was discarded early on by the Trump administration, a decision so damaging to American interests that it will always be a mystery for economic and political historians.

In any case, any careful analyst will find that the US government's claims about the positive attributes of the new agreement are as misguided as were their claims about the ills caused by NAFTA (Krueger, 2018). As a result of the US negotiators' pullback from their original demands, there will be a mechanism, if approved by the respective legislative branches, to keep markets open among the three partners but it will not be a better instrument than NAFTA for any of the three countries.

NAFTA negotiations aside, trade hostilities by the United States generally escalated significantly in 2018. In January, safeguard tariffs on solar panels and washing machines were announced. Next, invoking national-security arguments (section 232 of the Trade Expansion Act of 1962), an implausible argument for commodity metals, the US government imposed high tariffs on imports of steel and aluminum from China (effective in March) as well as the European Union, Japan, Turkey, Canada, and Mexico (effective early July 2018). Predictably, all the affected trade partners responded at once by announcing their own retaliatory trade actions.

The confrontation with China intensified with the announcement (effective in early July 2018) of tariffs on US imports from that country worth \$34 billion. The stated rationale was unfair trade practices (under section 301 of the Trade Act of 1974). By September 2018, the total value of Chinese imports subject to US section 301 tariffs had risen to \$250 billion, with tariffs on a further \$236 billion threatened.

It did not take long, in fact only a few hours, for China to respond in kind to the US action. At the time of writing, the Trump administration is vowing to react with even more tariffs on imports from countries challenging its arbitrary actions.

The trade aggressiveness, rather than an intelligent use of diplomacy, against China is difficult to understand, not only because almost no hard evidence has been provided about the imputed unfairness of China's own trade practices, that if true would warrant a strong



case to be judged by the World Trade Organization (WTO) appellate body, but also because it seems to ignore other aspects of the already significant interdependence between the American and Chinese economies. Among other things, the US government overlooks the effect of China's imports from the US in supporting the latter's economic growth as well as the favorable impact of the lower price of Chinese exports on the real wage of American workers. It equally seems to dismiss that the US trade deficit with China helps to feed the latter's current account surplus which is a simple consequence of the excess savings available in the Chinese economy and that the US has been happy to borrow over many years to compensate for its own very low savings rate.

It is hard to know whether the American administration really believes that sooner rather than later China and the other targeted countries will succumb to the United States' outlandish demands, and thus deliver Mr. Trump a "win" in the still incipient confrontation. If this were the assumption—most likely a wrong one—the trade war could reach epic proportions, with rather irreversible damage. Even worse, however, the US authorities could be envisioning a scenario in which the affected parties implement full recourse to the WTO, and this is taken as an excuse to withdraw from that institution, as President Trump has sometimes threatened to do.

This episode of American neo-mercantilism can hardly have a happy ending, simply because it has been launched on very wrong premises and with questionable objectives. The US government's ongoing policy not only ignores the notion of comparative advantage and its modern incarnation into complex supply chains, but also the essential insight from open-economy macroeconomics that the difference between an economy's national income and its expenditure is what drives its current account and trade balances. Playing with trade policy without looking at the underlying variables of income and expenditure is bound to be futile and counterproductive. Furthermore, focusing on bilateral balances to fix the aggregate one makes the undertaking even more pointless.

The discussion about the NAFTA renegotiation and the other trade actions undertaken by the current American government are highly relevant to a key inquiry of this article: the future of globalization. As claimed above, US neo-mercantilism has the potential to cause enormous damage to the process of increasing global economic interdependence built during almost three quarters of a century. How far and how deep the newly adopted American protectionist stance is taken will determine, more than any other circumstance, whether modern globalization is in its twilight or simply recedes temporarily. Of course, other factors, which must be duly acknowledged, will be at play to determine the ultimate outcome, but the decisive weight of US policies need to be factored properly into any exercise of prognosis about globalization.

If the capricious withdrawal from the TPP, the arbitrary imposition of import tariffs against products of its main trading partners, and the unjustified rhetoric that accompanied the renegotiation of NAFTA were the guide to predict the gravity of US policies, it would be prudent to envision a dramatic compression of globalization in the years to come. This would happen in a scenario where a tit-for-tat vicious cycle of rising trade barriers happens along with the annihilation, formal or de facto, of the WTO and the other cooperative instruments that exist to govern international trade and investment. Obviously, this would be an outcome, economic and otherwise, of practically catastrophic proportions for the US, its main trading partners and all the other participants in the global economy.

A considerably less disruptive scenario could be imagined if the process to devolve NAFTA into a United States-Mexico-Canada agreement (USMCA) were the relevant guide. As argued before, the new deal (if ratified), while not really being a meaningful improvement



over NAFTA, would still be capable of allowing a reasonable degree of mutually convenient integration, not optimal but substantial, among the three partners. Fortunately for all involved, the protectionist rhetoric displayed by the American authorities was not matched at the end by their actions to conclude the deal. In fact, the termination of NAFTA would have been the only outcome matching the aggressive and bombastic positions they held from the beginning to right before the end of the talks. The equally bombastic exaggerations utilized to announce the purported virtues of the new agreement could also be suggestive. Demand unreasonably, negotiate aggressively, make any possible deal, declare victory, and move on, seemed to be the script used by the US government in the NAFTA negotiations. If this were truly the blueprint that the US government intends to follow to restructure the country's trade relations with the rest of the world, then the damage, if not negligible, will be contained. It might even be the case that as trade is disrupted by the US protectionist actions and its partners' retaliation, the damage in terms of jobs, output, and lower real wages could lead to a shift in the American position where rationality prevails over the wrongheaded populist instincts exhibited so far.

## **US neo-mercantilism has the potential to cause enormous damage to the process of increasing global economic interdependence built during almost three quarters of a century**

But even in the least pessimistic scenario, other important issues will have to be addressed in the years to come if a twilight of contemporary globalization is going to be avoided and allowed to continue being, on balance, a powerful force for prosperity and international peace and security. For one thing, all the other major economic powers of the world should deal intelligently with the ongoing American predicament with a view to certainly containing the aggressiveness of the US actions while doing their utmost to protect the rules-based international system. Those powers should commit their collective and coordinated action to compensate for the retrenchment by the US from the multilateral institutions and the provision of global public goods. They will have to be more proactive in their support of institutions and agreements such as the United Nations, the Bretton Woods institutions, the Paris Agreement on Climate Change, and many more. Although it will be very hard, if not impossible, to do it without the concurrence of the United States, the reform and strengthening of the multilateral system, not only in its purely economic aspects but in its entirety, is a necessary condition to deal effectively with the challenges posed by economic interdependence while benefiting the most from it. On the trade front, it is necessary and unavoidable, although unfortunate, that they continue reacting in a targeted fashion to the US trade restrictive moves, but it must be done in a way that complies with the WTO framework. They must also be more forthcoming about their support for that institution, abandoning the passivity or even free-riding attitude that has sometimes prevailed in the past, not least over the failed Doha Round.

Those powers will also have to address more decisively some important domestic challenges. For example, China, whose leadership has come forth to champion globalization in the face of the American retreat, should, in its own interest, move faster in its process of domestic reform. The European Union, nowadays the most benign pole on the map of world powers, should expedite its consolidation, a task that, among many things, implies dealing properly with the march of folly that Brexit is and do what it takes to make the European Monetary Union (EMU) unquestionably viable and resilient.



Crucially, for globalization to deliver to its full potential, all governments should take more seriously the essential insight provided by economics that open markets need to be accompanied by policies that make their impact less disruptive and more beneficially inclusive for the population at large.

Advocates of globalization should also be more effective in contending with the conundrum posed by the fact that it has become pervasive, even for serious academics, to postulate almost mechanically a causal relationship between open markets and many social and economic ills while addressing only lightly at best, or simply ignoring, the determinant influence of domestic policies in such outcomes. This identification problem is adversely consequential for many reasons but mainly because it leads to bad, insufficient, or at best irrelevant policy prescriptions.

Linking abominable inequities to trade and even to technological progress, as has become fashionable, misses the point entirely on two important accounts. First, because it denies that those inequities are much more the consequence of explicit domestic policies unrelated to trade issues. By focusing on the latter, the important fact that those inequities are fundamentally the result of past political choices is overlooked. Second, by committing this omission, it becomes more likely to incur serious policy mistakes with the practical consequence that the inequities purported as undesirable will tend to be further perpetuated.

Trade policy will never be a first-best one to deal with the problems that rightly have increasingly captured the attention of citizens and political leaders in most countries, both developed and emerging, such as poverty, increasing inequality, and marginalization. These ills—less the result of historical initial conditions than of deliberate policy decisions by political leaderships unduly influenced over time by the holders of economic power—if they are to be addressed seriously, it must be done with institutions and policies conformed explicitly for such purposes (Zedillo, 2018). This is a clear-cut political choice that obviously poses a trade-off with other objectives, such as low taxation, that may be considered important by some economic and political constituencies. The real dilemmas must be acknowledged and acted upon, and not evaded as is done when tweaking trade policy is wrongly alleged to be the instrument to address unacceptable social ills.



## Select Bibliography

—Baldwin, Richard. 2014. "Trade and industrialization after globalization's second unbundling: How building and joining a supply chain are different and why it matters." In *Globalization in an Age of Crisis: Multilateral Economic Cooperation in the Twenty-First Century*, Robert C. Feenstra and Alan M. Taylor. Chicago: University of Chicago Press.

—Buitter, Willem H. and Ebrahim Rahbari. 2011. "Global growth generators: Moving beyond emerging markets and BRICs." *Citi Global Economics*, 21 February 2011.

—Global Trade Alert (<https://www.globaltradealert.org>).

—Hoekman, Bernard (ed.). 2015. *The Global Trade Slowdown: A New Normal?* A VoxEU.org eBook, Center for Economic and Policy Research (CEPR) Press.

—International Monetary Fund. *World Economic Outlook* various issues.

—International Monetary Fund. 2016. "Global trade: What's behind the slowdown?" Chapter 2 of the October 2016 World Economic Outlook, *Subdued Demand: Symptoms and Remedies*, Washington.

—Krueger, Anne O. 2018. "Trump's North American trade charade." *Project Syndicate*, October 12, 2018.

—Maddison, Angus. 2001. *The World Economy: A Millennial Perspective*. Development Centre Studies, Organisation for Economic Co-Operation and Development (OECD) Publishing.

—McKinsey Global Institute. 2017. *The New Dynamics of Financial Globalization*. McKinsey & Company, August 2017.

—Office of the United States Trade Representative. 2017. "Summary of objectives for the NAFTA renegotiation," July 17, 2017 <https://ustr.gov/sites/default/files/files/Press/Releases/NAFTAObjectives.pdf>

—Zedillo, Ernesto. 2017. "Don't blame Ricardo – take responsibility for domestic political choices." In *Cloth for Wine? The Relevance of Ricardo's Comparative Advantage in the 21st Century*, Simon Evenett (ed.). CEPR Press.

—Zedillo, Ernesto. 2018. "How Mexico and Canada saved NAFTA." *The Washington Post*, October 8, 2018. <https://www.washingtonpost.com/news/worldpost/wp/2018/10/08/nafta-2>.



**Victoria Robinson**  
University of York

Victoria Robinson is currently Professor of Sociology at the University of York, UK, and Director of the Centre for Women's Studies. She has also held academic posts at the universities of Sheffield, Manchester, and Newcastle. In the early 1990s she was responsible for setting up one of the first undergraduate degrees in Women's Studies in the UK. She has published books and articles in the areas of women's, gender, and masculinity studies, specifically in the areas of feminist theory, gender and sexualities, men and masculinities, fashion and footwear cultures, risk sports, and debates in women's, gender, and masculinity studies in the academy. She is currently co-editor, with Professor Diane Richardson, of Palgrave's international book series *Genders and Sexualities*.

Recommended book: *Introducing Gender and Women's Studies*, Victoria Robinson and Diane Richardson (eds.), Palgrave Macmillan, 2015 (4th edition).

**This article is concerned with the question of progress made on gender issues in a global context, specifically in terms of how far gender equality has been achieved, or not, in the past decade. It also reflects on how we might tackle one of the most pressing social, economic, and political issues of our times and effectively address this in the next decade and beyond. In so doing, it also considers the effects of political, social, and economic shifts on women's (but also men's) lives in both global and everyday contexts. In addition, how individuals and groups are resisting and challenging gender inequalities and attempting to intervene and correct the causes and consequences of gendered power imbalances will be discussed.**



To look at all areas of gendered life and inequality is beyond the scope of this piece. Therefore, I will discuss arguments that have been put forward that argue a case for the continuing existence of international gendered power relations in a number of specific areas: initially, education and violence. These arguments suggest that gendered inequality is visible in both public and private spheres, especially the economic, political, and social aspects, and provide evidence across some of the most pressing examples of gendered inequalities. The validity of the arguments that gender inequalities are still entrenched and persist over time, place, and culture will initially be compared to alternative claims that gendered power relations, and thus inequalities, are gradually being eroded. Moreover, given the current academic focus on the concept of intersectionality, that is, how the variables of class, sexuality, race, and ethnicity, for example, intersect in relation to people's gendered experiences, this concept is included in discussion here. The case study of women's global activism will provide a framework to further discuss these issues and take up some of the questions that have been raised.

In addition, I will conclude with an argument that the study of inequality in relation to gendered identities, relations, and experiences must continue with, and further utilize, the relatively recent exploration of the study of men and masculinities if the theoretical analysis of gender is to be enriched, and inform the (still) much-needed focus on women's experiences alone. I also argue the view that in relation to the future academic study of gender, as well as people's everyday gendered experiences in a global context, that to set the agenda for a more equal future society, we need to link gender much more closely to other inequalities, such as ethnicity and sexuality. I also consider forging new links between the academy and recent forms of activism in an international context.

There are those who argue that gender inequalities around the world are getting less. Dorius and Firebaugh, in their 2010 study, investigated global trends in gender inequality. Using data to research developments in gender inequality in recent decades across areas including the economy, political representation, education, and mortality, they conclude that a decline in gender inequalities can be seen which spans diverse religious and cultural traditions. Despite the fact that population growth is slowing this decline, as population growth is more prevalent in countries where there is most evidence of gender inequality. However, even optimistic studies such as this admit that:

Optimism about the future of global gender equality must be cautious for two reasons. First is the obvious point that there is no guarantee that current trends will continue. Second, gender equality can be seen as a two-step process that can be summarized colloquially as 'first get in the club, then attain equality within the club.' Most of the indicators we examine here focus on attaining membership in the 'club'—enrolling in school, joining the economically active population, becoming a member of the national legislature. Gender parity on these indicators is only part of the story since, to cite one example, men and women are entering highly sex segregated labor markets, at least in industrialized countries (Charles and Grusky, 2004). (Dorius and Firebaugh, 2010: 1959).

There is overwhelming evidence that would refute this and other similar linear perspective accounts of progress in gender matters. The recent *World Inequality Report* (WIR2018; Avaredo et al., 2018) is a major systemic assessment of globalization outlining income and wealth inequality, and documents a steep rise in global economic inequality since the 1980s, and this is despite strong growth in many emerging economies. It is within this context that any

analysis of gendered inequalities must be placed. Of course, poor men, men of color, gay men, to name just some of the groups other than women, are affected by economic, racial, and sexual discrimination. But overall, it is women who bear the brunt of poverty, violence, and inequality in the workforce, for example. Indeed, on average, the world's women earn twenty-four percent less than men (UNWomen, 2015).



## **In relation to the future academic study of gender, as well as people's everyday gendered experiences in a global context, to set the agenda for a more equal future society, we need to link gender much more closely to other inequalities, such as ethnicity and sexuality**

Discussing her recent book (Campbell, 2014a), UK-based writer and journalist Beatrix Campbell (2014b) takes the stance that such liberal thinkers have an over optimistic view that the road to gender equality is now within sight. Conversely, she argues this is largely an illusion. She defines the current era as one of “neopatriarchy” where rape, sex trafficking, and the unwon fight for equal pay characterize societies. Earlier, in 2014c, she forcefully argued that in the first decade of this century, the actual conditions which she deems necessary to end inequalities between the sexes have, in fact, been extinguished:

In this perceived era of gender equality, there is a new articulation of male social power and privilege. There is no evolutionary trek toward equality, peace and prosperity. The new world order is neither neutral nor innocent about sexism: it modernises it. Masculinities and femininities are being made and remade as polarised species. (Campbell, 2014b, c: 4).

Certainly, there is much available evidence to support Campbell's view. As Boffey (2017) reports regarding the latest EU gender equality league table, there has only been slow progress in relation to gender equality across Europe in the years between 2005 and 2015. He notes that the overall score for gender equality (when a matrix of data is taken into account) only grew by four points, to 66.2 out of 100, with 100 signifying absolute gender equality. Further, he reports that:

The gender gap in employment in the EU is ‘wide and persistent’, the index report says, with the full-time equivalent employment rate of 40% for women and 56% for men. Income gaps have narrowed, but on average women still earn 20% less than men, and the average masks huge disparities across the EU. (Boffey, 2017: 6).

In addition, the data reveals that for every third man in the EU who does daily housework and food preparation, this contrasts to eight in ten women who undertake the same tasks. And in the private sphere, nearly every second working woman has at least an hour of work with childcare, or other caring duties, contrasted with around about a third of working men. As I go on to document, extensive evidence also exists for the persistence of gender inequality outside of the EU more globally, in and across both public and private spheres and across multiple sites. However, it is important to note that this evidence can be interpreted in different ways. By highlighting two key substantive areas of education and violence, which have



been sites of gendered inequality focused on over the last decade by policy makers, activists, and academic makers alike, it can be seen that discussion has ranged between a narrative of progress, to varying or lesser degrees, or a more pessimistic viewpoint. How to escape this often dichotomous position is something that needs our attention.



## Education

In the last decade, the narrative of Malala Yousafzai, the Pakistani activist for female education who was awarded the Nobel Peace Prize in 2014 for her brave resistance against the Taliban doctrine (which had effectively banned girls from attending school and thus the human right to an education) has become well known. In a sense, this example can be seen as symbolic of two tendencies within current thinking. One tends to either read Malala's heroic efforts as evidence that demonstrates the much-needed necessity of campaigns for girls' education, especially in countries where they are denied equal access as boys may have. Or, as Rahman (2014) argues, her iconic, global status afforded by the media, celebrities, and governments is actually very problematic, masking as it does the continued educational inequalities which have their roots in complex historical, geopolitical, and development aspects in an Internet age.

Certainly, it is undeniable that a number of factors still exist worldwide that prevent girls from access to schooling due to issues, for instance, of girls leaving education on becoming child brides in countries such as Zambia, the sexual harassment and violence girls face in countries like South Africa, and the impact of war on girls' education in places like Rwanda or the Sudan (Ringrose and Epstein, 2015). Clearly, these issues are complex and vary across time and geographical location, but, even in the Global North, gendered inequalities in education still endure.

One current example of this is the recent revelation that, in 2018, Tokyo Medical University marked down the test scores of young women applying to embark on a career in medicine, to ensure more men became doctors. The university had, apparently, systematically kept the ratio of female students to around a third. The reason given that the authorities were concerned with their ability to want to continue working after starting a family. Such examples reveal the sustained, and hidden, institutional sexism in education that both serves to exclude young women from reaching their full potential and eventually affects their future earning potential. It also reflects how continuing societal assumptions, based on stereotyped and biologically essentialist notions of gender, still have purchase across the world.

## Violence

Another issue which has been highlighted by both scholars and activists is the enduring aspect of violence against women in its many manifestations. As Liz Kelly (2015) observes, every week there is a story in the media nationally, or internationally, featuring violence against women. Such violence includes Female Genital Mutilation (FGM), rape and assault, trafficking and honor-based violence, sexual violence in countries undergoing conflict, domestic violence, violence and the issues of migration, asylum seeking and the refugee crisis. The focus has also been on how the political responses to these diverse areas are gendered and impact on women's identity and relationships with others, as well as on the unreported, everyday acts of violence in both the Global South and the Global North. This is, however, a gender inequality



which has been much targeted by global activism to combat these diverse manifestations of violence and their unequal effects on women. In addition, in the last decade, there has been more of a recognition than hitherto that men, though in the majority of perpetrators of violence against women, in certain contexts and age groups also face violence, most notably from other men. A timely example of this is that currently, in South Africa, the highest cause of mortality among poor young black men is violence, including murder at the hands of other men, often linked to crime and gangster-related activities.

This more comprehensive approach to combating violence can be seen in the example of the existence of the International Day for the Elimination of Violence Against Women, in 2016, which was then followed by Sixteen Days of Activism Against Gender-Based Violence. What is particularly interesting in relation to this initiative was that the violence toward women was acknowledged and debated in the context of its impact on women, men, and children. Further, it was recognized that both women and men strive to help both victims and perpetrators, as well as challenge global violence in all its forms.

## **In 2018, Tokyo Medical University marked down the test scores of young women applying to embark on a career in medicine, to ensure more men became doctors**

In addition, academics are currently developing new methodologies to measure violence and make more visible the previously hidden extent of gender-based violence (Towers et al., 2017). It would have been unimaginable, even a decade ago, that in 2018 New Zealand would have passed legislation granting victims of domestic violence ten days' paid leave which will allow them to be able to leave their partners, protect themselves and their children, and seek out new housing.

### **New Forms of Women's Global Activism**

The case study of women's global activism raises further interesting and crucial questions which the discussion so far has started to address. It allows a new focus on continued and structural gendered power relations, discrimination, institutional and structural inequalities, and the impact of this on everyday lives, but also affords a discussion of people's agency, optimism, and collaboration, as well as the increasing role of social media in activist campaigns and academic analysis.

Women have, over the past decade, for example, been involved in the far-reaching Middle East revolutions, protests in Delhi and elsewhere in India over widespread incidents of rape and sexual assault, and the much-documented women's movement protests in the US over Donald Trump's policies. As Nickie Charles (2015) notes, this resurgence of feminist action has partly been understood as the continuation of "third wave" feminism, with a history at least stretching back to the suffragettes in the UK and the Women's Liberation movement worldwide from the 1970s. Others, however, have viewed such renewed international activism as heralding a new era of protest, heralded by social media and its potential to make such protests for gender inequality truly global, in ways which were not possible before. In addition, many men of a younger generation have no hesitation in calling themselves feminists and



**Education and violence  
have been sites of gendered  
inequality especially focused  
on over the last decade by  
international activism**

A man walks through a group of women  
participating in the performance "Women in Black  
Acción" created by artists May Serrano and María  
Seco to protest against gender-based violence.  
November 19, 2015, Malaga, Spain









working with women on a range of issues and campaigns. The LGBTQ (lesbian, gay, bisexual, transgender, and queer) movement has allowed traditional ideas of only two genders existing to be problematized by a reconceptualization of the concept of gender and claims to gender fluidity. Further, the increasing acceptance of transgendered people (though not in all parts of the world and not without much debate and controversy in terms of who is able to call themselves a woman or man, depending on the resulting arguments around the sex assigned at birth) has been a key and continuing issue over the last and future decade (see Jackson and Scott, 2017). Lastly, the emphasis placed on intersectionality and how gender links to other categories, such as race and ethnicity, age and class, informs current campaigns and continues to be a central concern of feminists and the women's movement.

## **Many men of a younger generation have no hesitation in calling themselves feminists and working with women on a range of issues and campaigns. The LGBTQ movement has allowed traditional ideas of only two genders existing to be problematized by a reconceptualization of the concept of gender and claims to gender fluidity**

The #MeToo online campaign, which followed in the wake of the sexual misconduct campaigns against the Hollywood producer Harvey Weinstein, in 2017, drew attention to the sexual assault and sexual harassment of women, initially in Hollywood and the entertainment business. As Lawton (2017) notes, on Facebook, the comments and reactions to the campaign totaled more than twelve million in twenty-four hours. Moreover, Alka Kurian (2018), in *The Conversation*, reflects on the #MeToo movement, arguing that current legal and moral interpretations of "consent" are no longer fit for purpose, especially for a younger generation of media-adept women and men, who are questioning traditional gender and sexual roles and identities. She further notes the global potential of social media and similar online campaigns:

In the early 21st century, millennial Indian women launched a radically new kind of feminist politics that had not been seen before. Inspired by a vocabulary of rights and modes of protest used by the youth across the world, such as Occupy Wall Street and the Arab Spring, they initiated a series of social media campaigns against the culture of sexual violence. (Kurian, 2018: 4).

Such campaigns are not without their critics; for example, there are diverse views on what can, or cannot, be defined as sexual abuse, and the problem of establishing consent. Nor can it be assumed that such campaigns have the same affect globally. Japan, for example, has recently been highlighted as a place where girls and children are overrepresented in pornography, there is a rising sex crime rate, and the treatment of rape victims/survivors has been criticized. Evidence, some would argue, that such campaigns as the #MeToo movement cannot, by themselves, fully deal with structural inequalities and gendered power relations in capitalist societies. Some commentators also argue that online campaigning effectively takes the focus off the anticapitalism struggle. Moreover, even when global change is occurring, for instance, with Taiwan being poised to become the first country in Asia to legalize same-sex marriage, due to the efforts of the LGBTQ movement there, conservative groups, including



many Christian churches, are attempting to resist marriage equality, before the May 2019 deadline by which same-sex marriage will automatically become law.

Yet, to get even this far, Ting-Fang (2017, para 11) notes, on the efforts of Taiwanese activists, that: “This civil revolution is happening not only in the meeting rooms of the Legislative Yuan, but indeed, also on the streets and around the dinner table,” revealing the need for activists to engage with the public imagination in new ways and on diverse fronts.

Similarly, Utsa Mukherjee (2018) notes a watershed moment for India but also for the global queer rights movement, given the current move of the Supreme Court of India in decriminalizing homosexual acts between consenting adults. But also importantly points out that resistance to this law is as “old as the law itself,” and that the legal fight against such outdated colonial-era law started many decades ago as a protest against colonial marginalizing of “non-normative sexualities and gender expressions,” forcing such sexualities into Western categories and in the process criminalizing them. Such historical reflection reveals the need to acknowledge people’s experiences and protests in earlier decades, before the existence of social media and recent online campaigns, which can also reveal different priorities, as well as diverse ways of organizing.

Another example of how both activism and technology are changing peoples’ responses and ways of protesting against gender inequalities in its many forms is in relation to reproductive health, especially in respect of abortion rights:

Around Dublin, you might catch sight of a small sticker—on a lamp post, a wall, or inside the stall door of the women’s toilets—advertising ‘SAFE ABORTION WITH PILLS’ alongside a web address. The information on this sticker is technically illegal: it is advertising pills banned in Ireland and breaking Irish laws on the kind of information about abortion that can be legally distributed. The website advertised on this sticker will connect you with a global network of pro-choice volunteers who can advise you on how to access safe abortion pills, how to take them at home, and how to monitor your symptoms. (Calkin, 2018: 1).

In fact, the Republic of Ireland has now voted overwhelmingly to overturn the abortion ban in a referendum held in May 2018. Previously, abortion was only allowed when a woman’s life was at risk, though not in cases of incest, rape, or fatal fetal abnormality. Importantly, though, Calkin points out that forty percent of women globally reside in countries with highly restrictive abortion laws. Further, though only five countries ban abortion entirely, women worldwide face harsh restrictions and penalties when seeking an abortion. However, he contends that governments’ actions to control access to abortion is overall decreasing. A fact Calkin puts down to advanced communications and medical technology, but also, importantly, to cross-border transnational activists who seek to give alternate routes for women to access safe abortions.

If working across borders, the necessary existence of solidarity between genders, races, sexualities, classes, and ages of the actors involved in protesting and redressing gender equalities is essential for activists to be able to tackle increasingly complex and inter-related globalized issues. Pelin Dincer’s (2017) argument highlighting the question of women’s solidarity and working across differences is important to note when considering the effectiveness of international feminist movements, both theoretically and in activist terms. Her specific interest is the current fragmentation of the women’s movement in Turkey, and she uses as a metaphor for this the example of women marching there in protest against Donald Trump’s inauguration as US president. In so doing, she considers that some of the concerns that trans people and others had of such protests need to be voiced and heard. However, she



Demonstration at the Stonewall Inn in New York's Greenwich Village—a landmark for the gay rights movement—organized on February 23, 2017, to call for the protection of transgender and gender non-conforming persons



concludes that we must work with and not against such differences, if protest on a global scale is to be politically effective. Therefore, to this end, we need both a macro awareness of changing political and economic contexts in tandem with a more micro analysis of diverse activist practices and movements.



## Conclusion

Based on the above argument and evidence put forward, my concluding contention is that, in going forward, we can usefully focus on three aspects to continue to address the global issue of gendered inequality in innovative and more fruitful ways. These are: to further the contemporary debate and emphasis on intersectionality in relation to gender inequality; to highlight the increasing academic focus on masculinity and gender relations and its relation to feminism; and to rethink activism and its connection with the academy and others involved, especially in the light of technological advances. Simon Willis (2014) argues that: “Inequality is an urgent and complex problem. It is deeply entrenched in all areas of life. It is pervasively defended and supported, even by those who it damages. To my mind inequality is the main roadblock in our journey toward social justice, and we need an innovative approach to uprooting it that won’t produce the same negligible incremental change we’ve seen in recent years” (Willis, 2014: 1).

## **To address the global issue of gendered inequality in more innovative ways, it would be useful to further the contemporary debate and emphasis on intersectionality in relation to gender inequality and to highlight the increasing academic focus on masculinity and gender relations and its link to feminism**

Further, he feels that to address the structural and institutional causes of inequality, one of the main factors for doing so is the recognition of many, interconnected inequalities, as well as having an openness to work with diverse kinds of partners in a variety of ways. In a similar vein, the LSE Commission on Gender, Inequality, and Power, in 2015, was chiefly concerned with examining persisting inequalities between women and men in the UK. A key question the report asked was just how interconnected are inequalities across different sites of social life. It is a positive sign that policy makers, academics, and activists are constantly thinking through the possibilities of an intersectional approach in different contexts, despite some of the complex issues this raises.

The study by feminists or pro-feminist men in universities across the world on men as gendered beings and the meaning and experience of masculinity is one of the most important intellectual developments over the last decade. The examples discussed here have revealed that men can be oppressors but also victims, as well as collaborators in feminist causes. A recognition of men’s economic, political, and social power, as well as the issues faced by poor men and those of diverse races and ethnicities, for instance, can aid in a comprehensive picture of gendered inequality interacting with race and class, to name but two other facets of inequality. Thus, a more relational perspective on gender and inequality, while keeping in mind that women still bear the brunt of economic and other disadvantages, is important to develop further.





Lastly, as I have been writing this piece, the disturbing news has surfaced that the Hungarian government proposes to ban Gender Studies at universities in the country, at the start of the 2019 academic year. This is ostensibly because it was argued that employers were expressing no interest in employing a dwindling number of graduates of the subject and so the field is not seen as an economically viable one. Critics of such unprecedented state intervention and censorship of academic matters in Hungary have argued that, in reality, the ban is due to opposition to the government's conservative ideologies and policies. Since then, protests have ensued both in the streets and online. Further, the international academic community has joined together to oppose such sanctions and defend academic freedom for the unobstructed study of gender and gender inequalities in all its forms. Ann Kaloski-Naylor (2017) reminds us:

We need wider visions of resistance, ways out of the to and fro of arguments which seem to move us closer to disaster. This is what thinkers can offer, as well as our bodies and our posters on the streets and our ideas and petitions on the net. ...alternative visions that don't just respond to and recycle the immediate... (Kaloski-Naylor, 2017: 7).

If we are unable to even *think* about gender issues, it is of increasing importance that academics, practitioners, and activists continue to find new ways of speaking to each other on the issue of gender inequality. In so doing, as I have argued, the boundaries between academia and academics, civic and political institutions, and those who construct knowledge "outside" of such institutions, including activists in everyday life, have, by necessity, to become more fissured and blurred (Robinson, 2017).

## Select Bibliography

- Avaredo, F., Chancel, L., Piketty, T., Saez, E., Zucman, G. (eds.). 2018. *World Inequality Report*.
- Boffey, D. 2017. "UK gender inequality as bad as 10 years ago, EU league table shows." *The Guardian*, October 11, <https://www.theguardian.com/inequality/2017/oct/11/uk-no-further-forward-on-tackling-gender-inequality-eu-league-table-shows>.
- Calkin, S. 2018. "Abortion access is changing through technology and activism." *Discover Society*, <https://discoversociety.org/?s=calkin>.
- Campbell, B. 2014a. *End of Equality (Manifestos for the 21st Century)*. UK: Seagull Books.
- Campbell, B. 2014b. "End of equality." <http://www.beatrixcampbell.co.uk/books/end-of-equality-manifestos-for-the-21st-century>.
- Campbell, B. 2014c. "Why we need a new women's revolution." *The Guardian*, May 25, <https://www.theguardian.com/commentisfree/2014/may/25/we-need-new-womens-revolution>.
- Charles, N. 2015. "Feminist politics: From activism to representation." In *Introducing Gender and Women's Studies*, V. Robinson and D. Richardson (eds.). London: Palgrave Macmillan (4th edition).
- Dincer, P. 2017. "A feminist's fieldwork notes on women's solidarity and differences in Turkey." *Discover Society* 42, <https://discoversociety.org/2017/03/01/a-feminists-fieldwork-notes-on-womens-solidarity-and-differences-in-turkey>.
- Dorius, S. F., and Firebaugh, G. 2010. "Global gender inequality." *Social Forces* 88(5): 1941–1968.
- Jackson, S., and Scott, S. 2017. "Focus: Trans and the contradictions of gender." *Discover Society* 45, <https://discoversociety.org/2017/06/06/focus-trans-and-the-contradictions-of-gender>.
- Kaloski-Naylor, A. 2017. "Viewpoint: From fear to hope, from protest to resistance." *Discover Society* 42, <https://discoversociety.org/2017/03/01/viewpoint-from-fear-to-hope-from-protest-to-resistance>.
- Kelly, L. 2015. "Violence against women." In *Introducing Gender and Women's Studies*, V. Robinson and D. Richardson (eds.). London: Palgrave Macmillan (4th edition).
- Kurian, A. 2018. "#MeToo is riding a new wave of feminism in India." *The Conversation*, February 1, <https://theconversation.com/metoo-is-riding-a-new-wave-of-feminism-in-india-89842>.
- Lawton, G. 2017. "#MeToo is here to stay. We must challenge all men about sexual harassment." *The Guardian*, October 28, <https://www.theguardian.com/lifeandstyle/2017/oct/28/metoo-hashtag-sexual-harassment-violence-challenge-campaign-women-men>.
- Mukherjee, U. 2018. "India decriminalizes homosexuality." *Discover Society*, <https://discoversociety.org/2018/09/10/india-decriminalises-homosexuality>.
- Rahman, F. 2014. <https://blogs.state.gov/stories/2013/07/10/malala-day-promoting-education-all>: pp. 163–164.
- Ringrose, J., and Epstein, D. 2015. "Postfeminist educational media panics and the problem/promise of 'successful girls.'" In *Introducing Gender and Women's Studies*, V. Robinson and D. Richardson (eds.). London: Palgrave Macmillan (4th edition).
- Robinson, V. 2017. "Focus: Feminism in the academy and beyond." *Discover Society* 42, <https://discoversociety.org/2017/03/01/focus-feminism-in-the-academy-and-beyond>.
- Ting-Fang, C. 2017. "On the frontline: Marriage equality in Taiwan." *Discover Society* 42, <https://discoversociety.org/2017/03/01/on-the-frontline-marriage-equality-in-taiwan>.
- Towers, J., Walby, S., Balderston, S., Corradi, C., Francis, B., Heiskanen, M., Helweg-Larsen, K., Mergaert, L., Olive, P., Palmer, E., Stöckl, H., Strid, S. 2017. *The Concept and Measurement of Violence against Women and Men*. Bristol: Policy Press.
- UN Women. 2015. *Progress of the World's Women 2015–2016: Transforming Economies, Realizing Rights*. <http://asiapacific.unwomen.org/en/digital-library/publications/2015/04/progress-of-the-world-s-women-2015-2016>.
- Willis, S. 2014. "Policy briefing: Tackling inequality on the road to a just society." *Discover Society* 15, <https://discoversociety.org/2014/12/01/policy-briefing-tackling-inequality-on-the-road-to-a-just-society>.



**Barry Eichengreen**  
The University of California,  
Berkeley

Barry Eichengreen is George C. Pardee and Helen N. Pardee Professor of Economics and Political Science at the University of California, Berkeley, Research Associate of the National Bureau of Economic Research, and Research Fellow of the Centre for Economic Policy Research. Professor Eichengreen is convener of the Bellagio Group of academics and officials. He is a member of the American Academy of Arts and Sciences (class of 1997) and has been a fellow of the Center for Advanced Study in the Behavioral Sciences (Palo Alto) and the Institute for Advanced Study (Berlin). His most recent book is *The Populist Temptation: Economic Grievance and Political Reaction in the Modern Era* (Oxford, 2018).

Recommended book: *The Great Convergence: Information Technology and the New Globalization*, Richard Baldwin, Harvard University, 2016.

The global financial crisis of 2008 was widely seen as heralding the death of globalization. But, to paraphrase a remark widely attributed to the American novelist Mark Twain, “reports of its demise were greatly exaggerated.” Now, however, over a decade after the crisis, globalization hangs in the balance, as President Trump threatens tariffs on imports from a variety of US trading partners and challenges the norms of late-twentieth-century multilateralism, and as nationalist politicians of different stripes ascend to power in a growing number of countries. This chapter asks whether these unilateralist, nationalist trends appearing to constitute a serious threat to globalization are a temporary aberration—and if so accounts for their incidence and timing—or whether they in fact herald the much-anticipated retreat from globalization.



The years from 2008 to 2018 were an eventful period for the global economy, but no one would call them transcendent. The advanced economies suffered their most serious economic and financial crisis since the Great Depression, while events in Greece and elsewhere in Europe threatened the very survival of the Euro Area. A disappointing recovery gave rise to concerns about secular stagnation, the idea that deficient demand combined with stagnant productivity growth doomed the advanced countries to chronic slow growth.<sup>1</sup> In contrast, emerging markets, led by but not limited to China, escaped the crisis largely unscathed. They continued to expand throughout the crisis and for much of the subsequent decade.

As a result, the global economy grew at a more than respectable average annual rate of 3.4 percent over the years from 2008 to 2018.<sup>2</sup> Global trade continued to rise: after a dip in 2009, exports and imports recovered and mostly held steady as a share of global GDP. The institutional framework governing the operation of the global economy—a World Trade Organization to mediate trade disputes, an International Monetary Fund to monitor imbalances, and a World Bank to provide development assistance to poor countries—remained firmly in place. That globalization and growth could survive the turbulence buffeting the world starting in 2008 seemingly testified to the solid foundations on which the twenty-first-century global economy rested.

It all came apart in the final years of the period. In its June 2016 referendum, the United Kingdom voted to leave the European Union. In 2017 one of the first acts of the newly elected US president, Donald Trump, was to withdraw from the Trans-Pacific Partnership. The Trump administration declined to confirm the appointment of new members to the WTO's dispute settlement panel and in 2018 slapped tariffs on imports from China, Europe, and even Canada, provoking tit-for-tat retaliation. Cross-border investment flows were discouraged by high-profile government interventions, such as Beijing's refusal to approve Qualcomm's acquisition of Chinese semiconductor producer NXP, and Berlin thwarting China's acquisition of the German electricity transmission firm 50Hertz. The Chinese economy showed signs of slowing, and emerging markets from Argentina to Turkey experienced strains as the US Federal Reserve hiked interest rates. The stability of the global economy, it appeared, hung in the balance.

This sequence of events raises two questions, one about the past and one about the future. First, why was the reaction against the earlier globalization trend delayed by roughly a decade? In 2008–09 the advanced economies suffered the most serious downturn in eighty years, as noted. Export-dependent emerging markets experienced serious dislocations when advanced-country central banks responded with quantitative easing, pushing down their exchange rates in what were critically referred to as “currency wars.”<sup>3</sup> Yet there was no wholesale repudiation of the international system that bequeathed these disconcerting results. Instead, G20 countries reaffirmed their commitment to free and open trade and avoided beggar-thy-neighbor policies. Central banks provided one another with exceptional swap lines and credits. Governments sought to coordinate their fiscal-policy responses to the downturn. They expanded the resources of the International Monetary Fund to better equip it to meet the challenge of the crisis. Only after a delay of eight or so years, starting roughly in 2016, did the anti-globalization reaction set in with the Brexit referendum, the election of Donald Trump, and the rise of nationalist, anti-EU politicians and parties in a variety of European countries. What, in other words, explains this peculiar timing?

Second, what does this sequence of events imply for the future of the global economy? Is the negative reaction starting in 2016 an aberration? Is the election of an economic nationalist as US president the result of idiosyncratic factors—the political liabilities of his Democratic opponent Hillary Clinton and the untimely intervention of FBI director James Comey—and



a reflection of popular dissatisfaction with the public-policy response to the crisis, dissatisfaction that will dissolve now that growth has accelerated, unemployment has fallen, and wages have begun to rise? Similarly, is the populist turn in Europe simply the result of a one-of-a-kind Greek crisis and a temporary surge of refugees that has now receded? Or does the rise of economic nationalism in the West herald a fundamental rejection of the foundations of the global economy? And if so, what new system will take their place?



**In 2008–09 the advanced economies suffered the most serious downturn in eighty years, as noted. Export-dependent emerging markets experienced serious dislocations when advanced-country central banks responded with quantitative easing, pushing down their exchange rates in what were critically referred to as “currency wars”**

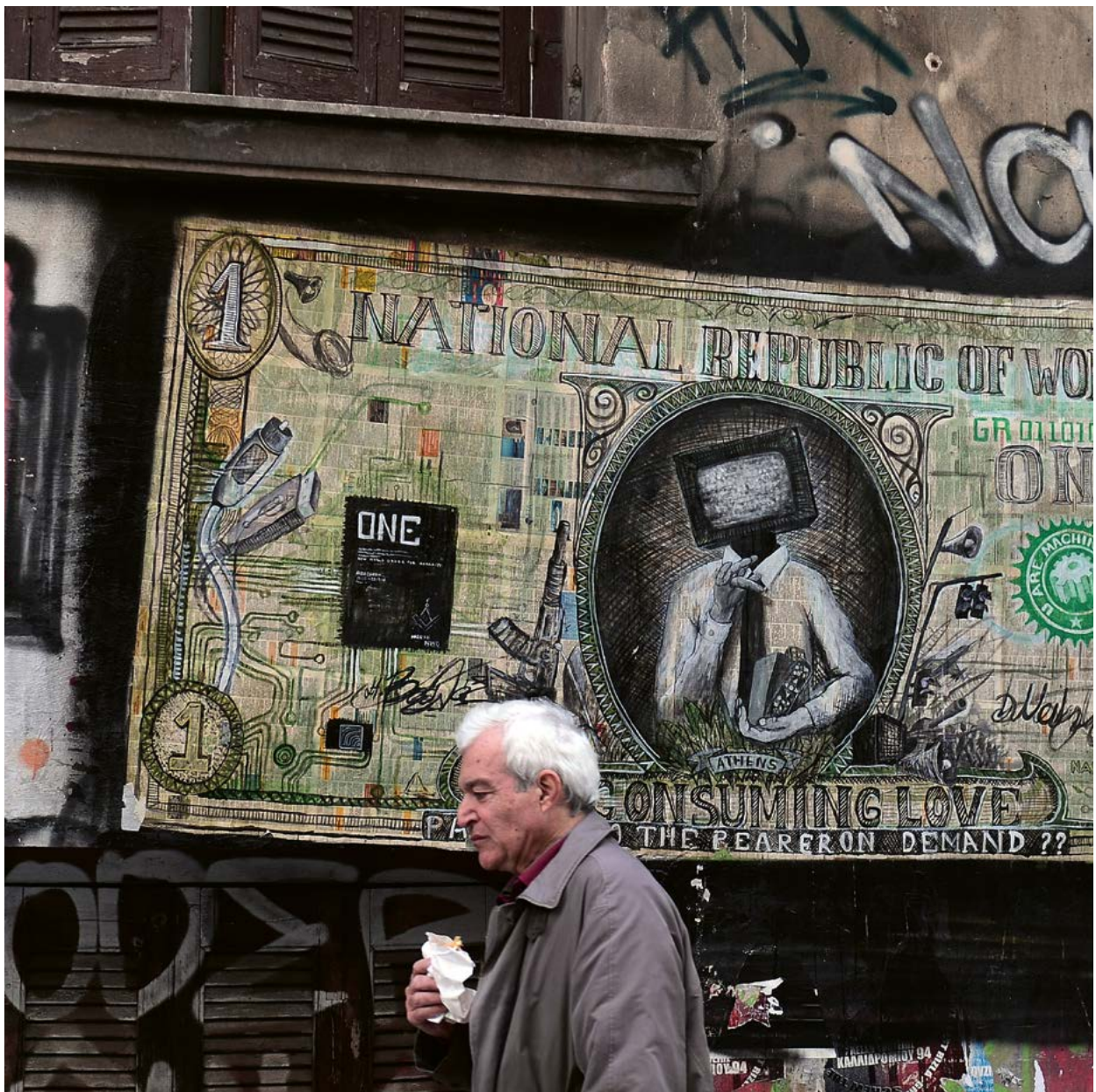
Formulating answers to these questions requires stepping back and contemplating the origins of the 2008–09 crisis that constitutes the backdrop to subsequent events. Like any complex episode, the crisis had more than one cause and, correspondingly, admits of more than one interpretation. My preferred interpretation runs as follows.<sup>4</sup> The crisis was a product, first and foremost, of inadequate financial supervision and regulation. In the US, the deregulatory movement had been underway for some years, fueled by free-market ideology and the political influence of large financial institutions; there had been no major banking and financial crisis in the US in the second half of the twentieth century to slow deregulation.<sup>5</sup> The final elimination of the Glass-Steagall Act separating investment and commercial banking in 1999 was only the symbolic capstone of these ongoing trends.<sup>6</sup> The inadequacies of light-touch regulation were further accentuated by the fragmentation of bank supervision across as many as six US government agencies and the absence of any agency whatsoever to oversee the operation of the so-called shadow banking system.

In Europe, completion of the Single Market and then the advent of the euro increased competitive pressure on the banks and encouraged regulators to lighten regulatory burdens in order to give national champions a leg up. European banks resorted to wholesale funding, lent aggressively to property developers, and loaded up on US subprime securities in the scramble to increase or at least maintain market share. Like in the US, the fact that Europe lacked a single bank supervisor—instead, it had a score of separate national supervisors who oversaw and, in practice, championed their respective national banks—meant that no one fully took into account the cross-border repercussions of their supervisory decisions.<sup>7</sup>

Second, the apparent stability of the macroeconomic environment encouraged risk-taking. Starting in the late 1980s, business-cycle volatility appeared to decline in the US and more widely. Whether the change reflected good luck—an absence of oil and commodity market shocks—or good policy—the shift to inflation targeting by a growing list of central banks—is disputed. But whatever its source, this so-called Great Moderation encouraged risk-taking by investors and by banks in particular.<sup>8</sup> It encouraged them to believe that business-cycle downturns were now milder, meaning that defaults would be fewer and making it possible to safely employ additional leverage and maintain slimmer capital and liquidity buffers. In this way the earlier period of stability set the stage for instability.



A man walking past graffiti in central Athens on February 4, 2015, following elections won by the Radical Left Coalition (Syriza) led by Alexis Tsipras





Third, loose monetary policy added fuel to the fire. Low interest rates encouraged investors to stretch for yield, governments to borrow, and households to load up on debt in what proved to be imprudent ways. The Federal Reserve, anticipating a serious recession, cut interest rates sharply following the September 11, 2001, attacks on the Twin Towers. When a recession of the anticipated severity failed to materialize, the central bank took time to normalize monetary policy. As a result, interest rates remained below the policy benchmark provided by the Taylor Rule.<sup>9</sup> The European Central Bank (ECB) appropriately set interest rates with the whole of the Euro Area in mind, but the resulting policies were uncomfortably loose for booming southern European economies on the receiving end of a tsunami of capital inflows.<sup>10</sup> The result was a securitization boom in the United States, housing bubbles in Ireland and Spain, and a government borrowing binge in Greece, which all came to grief when the US economy turned down.

## **In Europe, completion of the Single Market and then the advent of the euro increased competitive pressure on the banks and encouraged regulators to lighten regulatory burdens in order to give national champions a leg up**

Last in relative importance, but by no means insignificant, were global imbalances. Those imbalances, as measured by the absolute value of the current account balances of both surplus and deficit countries, reached their apex in 2006. The United States accounted for roughly two thirds of that year's cumulative deficit, while global surpluses were divided, roughly into thirds, between China and emerging Asia, the oil exporting countries, and Germany and Japan. By so willingly financing the US deficit, these surplus countries ensured that US interest rates would remain at lower levels than would have been the case otherwise, again encouraging investors there to stretch for yield and accept additional risk.

The stage was thereby set for the crisis. When first the US housing market and then the US economy turned down, bank and shadow bank balance sheets were impaired. And not just in the United States, since European banks had lent heavily to the property market while loading up on US subprime securities. Distress among subprime-linked hedge funds and investment banks, capped by the failure of Lehman Brothers in September 2008, led to panic and the seizing up of financial markets on both sides of the Atlantic. By the end of 2008 it was clear that a very serious recession was underway.

There are many valid grounds on which to criticize the subsequent policy response. After rescuing the investment bank Bear Stearns, the Federal Reserve aggravated the crisis by allowing Lehman Brothers to fail.<sup>11</sup> The fiscal-policy measures adopted in response were underpowered. While the headline numbers were impressive, chronic high unemployment and disappointingly slow recovery betrayed their inadequacy. President Barack Obama's economic advisors had advocated a larger fiscal package, but his political advisors counseled against.<sup>12</sup> European countries were even more phobic about budget deficits; the UK famously shifted into austerity mode already in 2010. European countries were reluctant to acknowledge, much less address, pervasive weaknesses in their banking systems. When the Greek crisis





erupted, they refused to contemplate debt restructuring, fearing for the stability of French and German banks. The European Central Bank (ECB) was slow to recognize the deflationary threat; instead of cutting, it raised interest rates in 2008 and again in 2011. Financial reform was inadequate: the Dodd-Frank Consumer Wall Street Reform and Consumer Protection Act of 2010 was weak soup by the standards of the Glass-Steagall Act of 1933, and European financial reformers did even less.

Yet the steps taken were enough for global growth to resume in 2010. GDP growth in the major advanced (G7) economies shifted from -3.8 percent in 2009 to +2.8 percent the following year.<sup>13</sup> In the advanced economies as a whole, it recovered from -3.4 percent to +3.1 percent. International cooperation was important for this success at staunching the bleeding. G20 governments agreed to coordinate their fiscal-stimulus measures at their February 2009 London summit. The Federal Reserve and European central banks were in continuous contact, and in 2008 the Fed extended dollar swaps not only to advanced-country central banks but also to those of four emerging markets: Mexico, Brazil, Singapore, and South Korea. The ECB for its part provided euro swaps. Governments foreswore overtly protectionist trade policies; average global tariff rates ticked up very slightly in 2009–10 but then resumed their trend decline, while the number of active trade disputes litigated at the WTO was less in every year between 2009 and 2013 than in the three years before the crisis.<sup>14</sup> Resort to capital controls was limited to the most severe crisis cases: Iceland, Greece, and Cyprus.

## **When the Greek crisis erupted, European countries refused to contemplate debt restructuring, fearing for the stability of French and German banks. The ECB was slow to recognize the deflationary threat; instead of cutting, it raised interest rates in 2008 and again in 2011**

In sum, while the international system came under strain, it survived the crisis intact. Political leaders and their economic advisors evidently believed that an international regime that had delivered prosperity and growth prior to 2008 would continue to do so once the emergency was overcome.

Why then did the outlook change so dramatically starting in 2016, with the Brexit referendum, the election of Donald Trump, and the more general political backlash against the prevailing rules-based global order? To repeat what was said earlier, it is tempting to invoke special factors. British Prime Minister David Cameron failed to deliver on his campaign promise to limit immigration, and he miscalculated in seeing a referendum on EU membership as a way of uniting the Conservative party and solidifying its grip on power. Donald Trump was helped by a weak opponent and by Russian interference in the US election.

Yet geographic and demographic variations in support for both Brexit and Trump suggest that more systematic factors were at work. So, too, does support for political leaders with autocratic, nationalistic, anti-globalization tendencies in a range of other countries.<sup>15</sup>

To start, there was the fact that recovery in the advanced countries, where this political reaction was centered, was disappointingly slow. Recoveries following crises tend to be slower than recoveries following plain-vanilla recessions because the banking and financial system is impaired.<sup>16</sup> There was also the failure to provide more policy support and the tendency to prematurely withdraw such support as was provided. Moreover, the majority of the income





gains that did occur in both the US and UK accrued to the wealthy—to the so-called one percent. In 2015, real median household income as measured by the US Census Bureau was still nearly two percent below its 2007 peak and nearly three percent below its level at the end of the twentieth century. The tendency for income gains to accrue disproportionately to the wealthy was less pronounced in continental Europe than in the English-speaking countries, but it was evident there as well. It is not surprising that these trends provoked a popular reaction against import competition and immigration, which were seen as benefiting capital at the expense of labor and causing wage stagnation.

But were voters right in blaming China and immigrants for these developments? Working-class wages had begun to stagnate and inequality had begun to rise already well before the 2008–09 crisis. The increase in inequality in the United States, in fact, dates back to the 1970s. This is why the median earnings of prime-age working men, adjusted for inflation, could fall by four percent between 1970 and 2010, despite the fact that the economy as a whole was continuing to expand. And what was true of the United States was true also for a range of other advanced economies. This timing clearly predates the China shock, which coincides more or less with that country's accession to the WTO in 2001. It long predates the increase in immigration from less developed countries to Europe and the United States in the early twenty-first century.

The explanation lies in the primacy of skill-biased technical change in both working-class wage stagnation and rising inequality. The substitution of machinery for assembly line workers accelerated in the 1970s and 1980s. Maintaining this machinery required relatively high levels of literacy and numeracy; it required education and skill. This shift visibly affected the demand for more- and less-skilled workers and therefore their compensation. Whereas in 1965, American workers with college degrees earned just twenty-four percent more than high-school graduates, that gap widened to forty-seven percent in the mid-1980s and fifty-seven percent in the mid-1990s. The situation in Europe and Japan differed in extent but not in kind.

Working-class voters were displeased by the failure of their governments to do more to temper the effects. Aware of their inability to turn back the clock on technical progress, they found it easier to blame immigrants and Chinese workers for their plight. The role of the financial crisis and slow recovery that followed was to provide a focal point for their anger, directing it toward imports, immigrants, and mainstream politicians and parties. It catalyzed their unhappiness with the ruling elites and led them to seek to dismantle the prevailing international order, in the UK by voting for Brexit, in the US by voting for Trump, and in countries like Italy, Poland, and Hungary by voting for parties and leaders antagonistic to the European Union.

Thus, the populist reaction that erupted starting in 2016 was more than a delayed response to the 2008–09 financial crisis. In fact, it reflected ongoing income stagnation, rising inequality, and a heightened sense of economic insecurity with roots that can be traced back to the 1970s and 1980s. Anger against immigrants and imports may have been misdirected, but this did not prevent opportunistic politicians from capitalizing on it.

Another key development over the 2008–18 decade was the emergence of China as a leading power with geopolitical ambitions. China is already the world's number one exporter and will soon overtake the US as the country with the largest GDP. It is building aircraft carriers and asserting itself in the South China Sea. It is using its economic leverage to forge strategic links with countries in South Asia, Central Asia, Africa, and Latin America. It is the third largest foreign investor and the number one source of foreign investment for a growing number of



**In both the United Kingdom and the United States, the majority of the income gains accrued to the wealthy, the so-called one percent**

Guests at the Frederick Law Olmsted Foundation's lunch leaving the Conservatory Garden at New York's Central Park in May 2017. This annual event brings together the city's wealthiest women to raise funds for park maintenance







**In the United States the median earnings of prime-age working men, adjusted for inflation, fell as much as four percent between 1970 and 2010**

General Motors assembly worker John Martinez (L.) was forced to retire in April 2009, just months before the automobile giant went broke. His hopes for his family had been pinned to what were previously generous pensions and medical coverage



countries. It is using its Belt and Road Initiative to increase the connections and dependence on China of other countries, in the region and beyond. Not just the Belt and Road but also the Asia Infrastructure Investment Bank, the BRICS Bank, the Chiang Mai Initiative Multilateralization, the PBOC's renminbi swaps, and the designation of official renminbi clearing banks for foreign financial centers are all indications of the intention of Chinese leaders to shape the international system to their liking.

China's scope for asserting that influence is, if anything, enhanced by the Trump administration's "America First" policies. By withdrawing from the Trans-Pacific Partnership, the administration squandered an opportunity to more deeply integrate East Asia into America's economic sphere. By slapping tariffs on European steel and aluminum exports and threatening tariffs on European exports of motor vehicles, in both cases on spurious national-security grounds, it has jeopardized its ability to work with its European allies to reform WTO rules to address concerns about China's subsidies for state-owned enterprises and its treatment of intellectual property rights. By casting doubt over European access to US markets, it has encouraged the European Union to contemplate closer economic ties with China. By threatening to withdraw from the WTO, it has thrown into question the very survival of the rules-based global order.

Against this backdrop, one can imagine several different scenarios unfolding in the coming decade. First, Trump's attempt to undermine the rules-based global order could be a temporary aberration. Business organizations, such as the US Chamber of Commerce, oppose Trump's tariffs and his efforts to tear down the North American Free Trade Agreement and the WTO. Members of Trump's own party in the Congress remain committed to free trade. Though intimidated by the president, they understand that the United States has benefited from the rules-based international order.

## **Another key development over the 2008–18 decade was the emergence of China as a leading power with geopolitical ambitions. China is already the world's number one exporter and will soon overtake the US as the country with the largest GDP**

Moreover, if Trump fails to wrest significant concessions from US trade partners and if his tariffs hinder the growth of US business and raise the cost of imported consumer goods, then voters may connect the dots between the president's policies and their economic ills. Trump himself may grow weary of his imperial presidency and give way to a more conventional US leader. The US will then resume its role as a constructive member of the WTO, a signatory of the Paris Climate Accord, and a participant in a reactivated Trans-Pacific Partnership. To be sure, America's unilateralist interlude will have weakened its influence. Other countries will have come to see it as a less reliable partner. In the meantime they will have negotiated agreements among themselves designed to reduce their economic and security dependence on the United States. Still, the Trump "pause," if that is all that it is, will not have fundamentally reshaped the international economic order.

One must be an optimist to entertain this scenario. A majority of Republican voters, according to public opinion polls at the time of writing, support Trump's restrictive trade policies. The American public has shown little ability to connect the president's tariffs with their negative consequences. Trump himself is a master of political misdirection.





A Trump follower at the National Mall in Washington DC celebrates Donald Trump's inauguration as 45th president of the United States on January 20, 2017





More fundamentally, popular support for free and open trade requires policies that compensate the “losers” through retraining and relocation schemes, and Americans’ deep and abiding hostility to government poses a formidable obstacle to mounting such programs. An approach to campaign finance that concentrates influence in the hands of the wealthy means that tax and public spending policies designed to redistribute income to those who are left behind are singularly unlikely to gain traction. In addition, America’s ideology of market fundamentalism is conducive to forgetting the lessons of the financial crisis. This creates a real possibility that post-crisis reforms will be rolled back, making another crisis more likely and bringing about the further polarization of public opinion.<sup>17</sup>

## **Business organizations, such as the US Chamber of Commerce, oppose Trump’s tariffs and his efforts to tear down the North American Free Trade Agreement and the WTO. Members of Trump’s own party in the Congress remain committed to free trade**

All these are reasons to think that the United States is singularly vulnerable to the siren song of protectionism, as it has been throughout its history, aside from only the second half of the twentieth century. They are reasons to believe, in other words, that the country’s unilateralist turn could prove enduring.

While some of these same tendencies are also evident in Europe, there are reasons for thinking that the EU, by comparison, is more likely to remain committed to multilateralism and openness.<sup>18</sup> European countries, having experienced extremist capture in the 1920s and 1930s, reformed their electoral systems to make capture by political outsiders like Trump less likely.<sup>19</sup> Lacking Americans’ deep and abiding hostility to government, they are capable of mounting public programs that compensate the losers from free and open trade. Because their economies are smaller, in many cases very significantly smaller, than that of the United States, they understand that their prosperity is intrinsically linked to trade, both multilateral trade and the EU’s Single Market. Even the UK, where inequality is greatest and the reaction against the EU is most intense, remains committed to economic openness.

But observing that Europe is likely to remain committed to openness is not the same as saying that it is capable of exercising the leadership needed to mold the international economic order. Foreign- and security-policy leverage and economic-policy leverage go hand in hand. Witness the ability of a geopolitically powerful United States to shape the post-World War II international economic order. The EU, for its part, has not shown the capacity to mount a common foreign and security policy; different European countries have very different views of what this would entail. The share of military spending in GDP is lower in Europe than in both the US and China. The continent is demographically challenged and is therefore poised to grow slowly. As Europe comes to account for a declining share of global GDP, it will become correspondingly less able to determine the nature of international relations.

This leaves China as the obvious candidate to occupy the space vacated by the United States. As the leading trade partner and source of foreign investment for a growing number of countries, it already has some capacity to influence the shape of the international economic order. The question is what kind of order China has in mind.



The answer is not straightforward. China is committed to openness and export-led growth. As President Xi Jinping put it at Davos in January 2017, China is committed “to growing an open global economy.” In other words, Beijing will not obviously want to reshape the global trading regime in more restrictive directions.

But in other respects, globalization with Chinese characteristics will differ from globalization as we know it. Compared to other leading economies, China relies more on bilateral trade agreements and less on multilateral negotiating rounds. In 2002 China signed the China-ASEAN Comprehensive Economic Framework Agreement, and subsequently it signed bilateral free trade agreements with twelve countries around the world, with more in the works.<sup>20</sup> Insofar as China continues to emphasize bilateral agreements over multilateral negotiations, this implies a reduced role for the WTO.

The Chinese State Council has called for a trade strategy that is “based in China’s periphery, radiates along the Belt and Road, and faces the world.”<sup>21</sup> I read this as suggesting that it has in mind a hub-and-spoke trading system, where China is the hub and countries along its borders, or periphery, are the spokes. Other researchers have previously foreseen the emergence of a hub-and-spoke trading system in Asia, and possibly other hub-and-spoke systems centered on Europe and the United States.<sup>22</sup> Were China to exert more forceful leadership of the global trading system, this scenario becomes more likely. Again, the implication is a diminished role for the WTO.

## **China is committed to openness and export-led growth. As President Xi Jinping put it at Davos in January 2017, China is committed “to growing an open global economy.” Beijing will not obviously want to reshape the global trading regime in more restrictive directions**

Beijing may then wish to elaborate other China-centered regional arrangements to complement its commercial agreements and to substitute for multilateral institutions such as the IMF and World Bank. It has the Asian Infrastructure Investment Bank, headed by Jin Liqun, as an alternative to the World Bank. The PBOC has made \$500 billion of swap lines available to more than thirty central banks. In 2016 the state-run China Development Bank and Industrial and Commercial Bank of China, acting presumably on behalf of the PBOC, provided Pakistan with \$900 million of emergency assistance to help it stave off a currency crisis. China’s regional approach may also be motivated by the fact that it is underrepresented, in terms of quota and voting shares, at the IMF, leaving the United States as the only country with veto power in the Fund. Were the Trump administration to block IMF reform by rejecting proposals for quota revision, or were it to withdraw from the New Arrangements to Borrow (NAB), which provide the IMF with a substantial fraction of its funding, there would be additional impetus for China to develop its regional alternative.<sup>23</sup>

A China-shaped international system may attach less weight to the protection of intellectual property rights, the appropriation of the intellectual property of multinational corporations by their Chinese joint-venture partners being a particular bone of contention between the Trump administration and the Xi government. Alternatively, one can imagine Beijing’s attitude on such matters changing as China itself becomes a developer of new technology. That said, the sanctity of private property, whether of residents or multinationals, has always been





A toy-factory worker stuffs a teddy bear with cotton at Wuhan, in China's Hubei Province







less in China's state socialist system than in Europe or the United States. Hence, intellectual property protections are apt to be less in a China-led global system.

More generally, China's government does more than that of the United States, through the provision of subsidies and instructions to state-owned enterprises and others, to shape the structure and evolution of its economy. Its so-called China 2025 Plan to promote the development of China's high-tech capabilities is only the latest instance of this general approach.<sup>24</sup> The WTO has rules intended to limit subsidies and to regulate the actions that countries can take to counter them. European Union competition policy is similarly designed to limit the use of subsidies and other state aids. A China-shaped trading system would lack, or at least limit, such disciplines.

A China-led international regime would also be less open to international investment. In 2017 China ranked behind only the Philippines and Saudi Arabia among the sixty plus countries rated by the OECD in terms of the restrictiveness of their inward Foreign Direct Investment (FDI) regime. One can think of these restrictions as another device giving Chinese enterprises space to develop their technological capabilities. This stance may change once China becomes a technological leader and as it becomes more concerned with outward than inward FDI. Or it may not. Similarly, China continues to exercise tight control over its financial system and maintains controls on capital inflows and outflows. While the IMF has evinced more sympathy for the use of such controls since the early 2000s, a China-led international regime would presumably be even more accommodating of governments that utilize them.

In sum, a China-led global economy would remain open to trade but be less multilateral, transparent, rules-based, and financially open than its Western-led predecessor.

## **China's government does more than that of the United States to shape the structure and evolution of its economy. Its so-called China 2025 Plan to promote the development of China's high-tech capabilities is only the latest instance of this general approach**

The past decade was a turbulent period for the global economy, opening as it did with the global financial crisis and closing with Brexit and the election of an American president with an "America First" agenda. The crisis brought to a head popular dissatisfaction with the prevailing economic order, although I have argued that it only catalyzed the already existing unease associated with ongoing trends: skill-biased technological change, rising income inequality, and the failure of governments to adequately equip individuals to cope with these problems. Of the various national reactions, that in the United States was most consequential, since it brought to office a Trump administration that essentially abrogated the country's leadership of the global system.

The implications for the future are far from clear. It could be that the Trump presidency is a passing phase after which the United States will reassert its support for and leadership of the multilateral system. But there are also reasons to doubt that this will be the case. The hostility of Americans toward government creates less scope than in, say, Europe for intervention to compensate the casualties of creative destruction and globalization. In a period of slow growth, unlike the third quarter of the twentieth century, this renders US support for and leadership of an open multilateral system problematic to say the least.

Whether America's unilateralist turn is temporary or permanent, there is more space either way for other powers to shape the future of the global economic order. Given its resources and ambitions, China is most likely to assume this role. A China-led system will remain open to trade, but it will be less open financially than the comparable US-led system. It will be organized more heavily along regional lines. It will admit of a more prominent role for the state. And it will be less rules-based and transparent. The idea behind admitting China to the WTO was that doing so would create pressure for the country to remake its politics and economics along Western lines. The irony is that the economic pressure that an internationally integrated China applied to the rest of the world, and to the US in particular, may end up having precisely the opposite effect.



## Notes

1. See Summers (2014).
2. This is the 2008–18 average annual change in gross domestic product at constant prices as given in the International Monetary Fund's World Economic Outlook.
3. A review of the debate is Saccomanni (2015).
4. This view is elaborated at greater length in Eichengreen (2015).
5. Some would point to the Savings & Loan crisis of the 1980s as an exception but, as Field (2017) shows, its macroeconomic impact was minimal.
6. A polemical but still very useful account of these trends is Johnson and Kwak (2010).
7. A good overview is Bayoumi (2017).
8. The most influential analysis of the roles of good luck and good policies in the Great Moderation is Stock and Watson (2003).
9. The argument is made most forcefully by Taylor (2015) himself.
10. Those capital flows and their determinants, including monetary policy, are analyzed by Lane (2013).
11. This is the forceful interpretation of Ball (2018). Others, such as Bernanke (2015), argue that the central bank had no choice but to allow Lehman to go under owing to its lack of eligible collateral, a conclusion that Ball vigorously disputes.
12. A popular account of the debate is Scheiber (2012).
13. Numbers are again from the IMF's World Economic Outlook data base.
14. Tariff rates are for sixty-four countries representing 91 percent of world trade in 2010, as in Nordhaus (2017). Figures for WTO trade disputes are from Azevedo (2014), figure 1.
15. The remainder of this section draws on Eichengreen (2018).
16. This point is famously if controversially made by Reinhart and Rogoff (2009).
17. Funke, Schularick, and Trebesch (2016) provide evidence that financial crises lead to further political polarization, and to swings to the political right in particular.
18. This is the conclusion of

Eichengreen (2018), where I make the arguments at more length.

19. I am thinking, for example, of the French electoral system, which allows the supporters of other candidates to unite behind the mainstream candidate in the second, run-off round of voting, or the German system, which requires a constructive vote of no confidence (agreement on a new leader) in order to dismiss a government.
20. Many of these bilateral agreements are with poor countries that are not particularly important as markets for Chinese exports. Rather, those agreements can be seen as a way for Beijing to exert additional economic leverage over the partner, and specifically to encourage it to adopt China's technological and product standards and manage their economies in the manner of China itself.
21. Tiezzi (2018), p. 47. The same terminology has been echoed in a series of subsequent official and semi-official publications.
22. See for example Baldwin (2009).
23. Activation of the \$255 billion NAB requires an 85 percent majority vote of NAB participants, giving the US a veto. In addition, the US will be required to withdraw from the NAB in 2022 absent legislative action.
24. See Baldwin (2018) for a discussion.

## Select Bibliography

- Azevedo, Roberto. 2014. "Azevedo say success of WTO dispute settlement brings urgent challenges." *World Trade Organization News*, at [https://www.wto.org/english/news\\_e/spra\\_e/spra32\\_e.htm](https://www.wto.org/english/news_e/spra_e/spra32_e.htm)
- Balding, Christopher. 2018. "Is China a market economy?" In *China: Champion of (Which) Globalisation?* Alessia Amighini (ed.). Milan: ISPI, 61–80.
- Baldwin, Richard. 2009. "The spoke trap: Hub and spoke bilateralism in East Asia." Swiss National Center of Competence in Research Working Paper no. 2009/28 (May).
- Ball, Lawrence. 2018. *The Fed and Lehman Brothers: Setting the Record Straight on a Financial Disaster*. New York: Cambridge University Press.
- Bayoumi, Tamim. 2017. *Unfinished Business: The Unexplored Causes of the Financial Crisis and the Lessons Yet to be Learned*. New Haven: Yale University Press.
- Bernanke, Ben. 2015. *The Courage to Act: A Memoir of a Crisis and its Aftermath*. New York: Norton.
- Eichengreen, Barry. 2015. *Hall of Mirrors: The Great Depression, the Great Recession and the Uses—and Misuses—of History*. New York: Oxford University Press.
- Eichengreen, Barry. 2018. *The Populist Temptation: Economic Grievance and Political Reaction in the Modern Era*. New York: Oxford University Press.
- Field, Alexander. 2017. "The macroeconomic significance of the savings and loan insolvencies." *Research in Economic History* 33: 65–113.
- Funke, Manuel, Schularick, Moritz, and Trebesch, Christoph. 2016. "Going to extremes: Politics after financial crises." *European Economic Review* 88: 227–260.
- Johnson, Simon, and Kwak, James. 2010. *13 Bankers: The Wall Street Takeover and the Next Financial Meltdown*. New York: Pantheon.
- Lane, Philip. 2013. "Capital flows in the Euro Area." *European Economy Economic Papers* 497 (April).
- Nordhaus, William. 2017. "The Trump doctrine on international trade." *VoxEU* (23 August), at <https://voxeu.org/article/trump-doctrine-international-trade-part-two>.
- Reinhart, Carmen, and Rogoff, Kenneth. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press.
- Saccomanni, Fabrizio. 2015. "Monetary spillovers? Boom and bust? Currency wars? The international monetary system strikes back." Dinner speech, BIS Special Governors' Meeting, Manila, 6 February, at <https://www.bis.org/publ/othp22.pdf>.
- Scheiber, Noam. 2012. *The Escape Artists: How Obama's Team Fumbled the Recovery*. New York: Simon and Schuster.
- Stock, James, and Watson, Mark. 2003. "Has the business cycle changed? Evidence and explanations." Jackson Hole Symposium of the Federal Reserve Bank of Kansas City.
- Summers, Lawrence. 2014. "Reflections on the 'new secular stagnation hypothesis'." In *Secular Stagnation: Facts Causes and Cures*, Coen Teulings and Richard Baldwin (eds.). London: Centre for Economic Policy Research.
- Taylor, John. 2015. "A monetary policy for the future." Remarks at an IMF conference on "Rethinking Macro Policy III, Progress or Confusion?" April 15.
- Tiezzi, Shannon. 2018. "Free Trade with Chinese Characteristics." In *China: Champion of (Which) Globalisation?* Alessia Amighini (ed.). Milan: ISPI, 39–60.



**Michelle Baddeley**  
University of South Australia

Michelle Baddeley is Director and Research Professor at the Institute for Choice, University of South Australia, and an Honorary Professor at the Institute for Global Prosperity, University College London. Previously, she was Professor of Economics and Finance at the Bartlett Faculty of the Built Environment, University College London, and before that was Fellow and Director of Studies (Economics) at Gonville & Caius College / Faculty of Economics, University of Cambridge. She has a Bachelor of Economics (First Class Honors) and BA (Psychology) from the University of Queensland, and a Masters/PhD (Economics) from the University of Cambridge. She specializes in behavioral economics, applied macroeconomics, labor economics, and development economics. She also has a keen interest in public policy and has collaborated with many public policy makers and government departments throughout her career. Her most recent books include *Behavioural Economics – A Very Short Introduction* (Oxford University Press, 2017), *Behavioural Economics and Finance* (Routledge, 2013/2018), and *Copycats and Contrarians: Why We Follow Others... and When We Don't* (Yale University Press, 2018).

Recommended book: *Behavioural Economics – A Very Short Introduction*, Michelle Baddeley, Oxford University Press, 2017.

**Consolidated by the award of the 2017 Economics Nobel Prize to behavioral economist Richard Thaler, behavioral economics is enjoying a golden age. It combines a diverse range of insights from across the social sciences—including economists’ powerful analytical tools alongside rich evidence about real human behavior from other social sciences—especially psychology and sociology. This article explores the evolution of behavioral economics and some key behavioral insights about incentives and motivations; social influences—including social learning, peer pressure, and group-think; heuristics and biases; decision-making under risk and uncertainty; present bias and procrastination; and nudging policy tools. These all illustrate how behavioral economics provides businesses and policy makers with a rich understanding of how real people think, choose, and decide.**



## Introduction



Today, it seems as though everyone is talking about behavioral economics. Governments are embedding behavioral insights into policy. Commercial businesses are using it to inform their marketing strategies. Lessons from behavioral economics are informing relationships between employers and employees. Even in the silos of academia, most applied research teams—most obviously other social scientists but also natural scientists, from neuroscientists through to behavioral ecologists, computer scientists and engineers—are keen to bring behavioral economists into the multidisciplinary teams so that they can connect their research with insights from behavioral economics. Why? Because behavioral economics combines a unique collection of insights from social science. It brings together economists' powerful analytical tools, traditionally applied in a restricted way to unraveling the economic incentives and motivations driving us all. But it also addresses the fundamental flaw in non-behavioral economics: its highly restrictive conception of rationality, based on assumptions of agents able easily to apply mathematical tools in identifying the best solutions for themselves or their businesses. Herbert Simon made some early progress in re-conceptualizing rationality in economics—via his concept of “bounded rationality”, that is, rationality bounded by constraints in information available or in cognitive processing ability (Simon, 1955). Modern behavioral economists have taken this further by bringing together rich insights from psychology to capture how economic incentives and motivations are changed, often fundamentally, by psychological influences. Neither the economics nor the psychology can stand alone. Without economics, the psychology lacks analytical structure and direction—especially in describing everyday decision-making. Without the psychology, economics lacks external consistency and intuitive appeal. Together, the subjects are uniquely insightful. Together, they enable us powerfully to understand what and how real people think, choose, and decide in ways that no single academic discipline has managed before—generating not only new theoretical insights but also new practical and policy insights that, at best, have the power to change livelihoods, prosperity, and well-being across a range of dimensions.

## The Past

Behavioral economics may seem to many observers to be a new thing, for better or worse. Most of the excitement about behavioral economics has bubbled-up in the past ten or so years. The first milestone was the award of the 2002 Nobel Prize jointly to economic psychologist Daniel Kahneman, alongside Vernon L. Smith—an experimental economist whose insights and tools inspired behavioral economists even though experimental economics is not behavioral economics. The second was the award of the 2017 Nobel to behavioral economist Richard Thaler, who has written colorfully about his contributions in his book *Misbehaving* (Thaler, 2106). Thaler is most famous for his work on behavioral finance and behavioral public policy—commonly known as “nudging,” named after his best-selling book with legal scholar Cass Sunstein—the book this year celebrating its tenth anniversary (Thaler and Sunstein, 2008). These thinkers have had enormous influence on modern policy—not least through advising the policy-making teams of the then US President Barak Obama and the then UK Prime Minister David Cameron. The establishment of a “nudge” unit in Cameron’s Cabinet Office spawned the growth of similar units around the world—from Australia to Lebanon to Mexico, to name just a few.



The progress of behavioral economics between the two milestones of the 2002 and 2017 Nobel Prizes mirrors the emergence of behavioral economics from a largely theoretical subject through to a subject that now has enormous real-world policy relevance—for public and commercial policy makers alike. It also has much to offer ordinary people in understanding some of the decision-making challenges they face. But behavioral economics is a much older discipline than these two twenty-first-century milestones might suggest. Some could argue that all economics should be about behavior if behavior is what drives choices and decision-making. Economics is the study of decisions after all. But, from the nineteenth century onward, economics started to move away from behavior as it might be richly understood in terms of the psychology of choice toward observed choices as a measure of revealed preferences. In providing a neat and simple story about these preferences revealed when we make our choices, the story can only be made sufficiently simple if economists assume that economic decision-makers are constrained by strict behavioral rules—specifically in assuming that consumers aim to maximize their satisfaction and businesses aim to maximize profits. In mainstream economics, consumers and firms are assumed to do this in the best way they can by implementing mathematical rules to identify the best solutions. Modern economists, in the process of building these neat mathematical models that captured these behavioral rules, stripped out all the socio-psychological complexities of real-world decision-making.

Historically, however, and before modern economics mathematicized the analysis of choice, economists spent plenty of time thinking about how the incentives and motivations that are the stuff of economic analysis are affected by psychological influences, including going all the way back to Adam Smith. Adam Smith is popularly associated with his defense of free markets in his 1776 masterpiece *The Nature and Causes of the Wealth of Nations*, in which he advocates that the “invisible hand” of the price mechanism should be allowed to operate without government intervention. But in his 1759 masterpiece—*The Theory of Moral Sentiments*—he also wrote extensively about sympathy and other social emotions that drive our interactions with others around us—key insights seen in modern behavioral economics research.

## The Present

We have said a lot about where behavioral economics comes from without saying too much about what behavioral economists actually *do*. In understanding this more deeply, we can look at a range of themes which behavioral economists explore to illustrate the power and relevance of their insights. Behavioral economics is now an enormous literature and doing justice to it all in one chapter is impossible, but a few key themes dominate and we will focus here on those insights from behavioral economics that are most powerful and enduring in illuminating real-world decision-making problems. These include behavioral analyses of incentives/motivations; social influences; heuristics, bias, and risk; time and planning; and impacts of personality and emotions on decision-making (see Baddeley, 2017 and 2018b, for detailed surveys of these and other behavioral economics literatures).

## Incentives and Motivations

As we noted above, economics is essentially about incentives and motivations—traditionally focusing on money as an incentive, for example in explaining a decision to work as a balanc-



ing act in which wages earned persuade workers to give up their leisure time. Psychologists bring a broader understanding of motivation into behavioral economics—specifically by disentangling extrinsic motivations from intrinsic motivations. Extrinsic motivations include all the rewards and punishments external to us—money is the most obvious, but physical punishments would be another example. Alongside these are intrinsic motivations—such as pride in a job done well, dutifulness, and intellectual engagement. Some of the most famous behavioral experiments were conducted by psychologist Dan Ariely and his team (see Ariely, 2008). Some of these experimental studies show that experimental participants’ decisions to contribute to a charity or public good are partly driven by social factors: people are more generous when their donations are revealed than when information about their donations is kept private. Intrinsic motivations drive effort, often as much and sometimes more than external monetary incentives. Students, for example, are often prepared to work harder for an intellectual challenge than for money. This illustrates that we are not driven just by external incentives and disincentives—whether these be money, physical rewards/punishments or social consequences. Chess, computer games, and also physical challenges associated with sport are all things that engage people’s attention and enthusiasm even without monetary rewards.

Disentangling intrinsic and extrinsic motivations is not straightforward, however. There is the added complication that extrinsic incentives “crowd out” intrinsic motivations. A classic study of this was conducted by behavioral economists Uri Gneezy and Aldo Rustichini. An Israeli nursery school was struggling with the problem of parents arriving late to pick up their children so they instituted a system of fines for latecomers. The fines had a perverse effect, however, in increasing the number of late pickups by parents rather than reducing them. Gneezy and Rustichini attributed this to a crowding-out problem: introducing the fine crowded-out parents’ incentives to be dutiful in arriving on time. Instead, parents were interpreting the fine as a price: in paying a fine they were paying for a service and so it became an economic exchange in which dutifulness in arriving punctually became less relevant (Gneezy and Rustichini, 2000).

A specific set of motivations that behavioral economists have spent a lot of time exploring are the social motivations and these are illustrated most extensively in what is possibly the most famous behavioral experimental game: the Ultimatum Game, devised by Werner Güth and colleagues (Güth et al., 1982). In the Ultimatum Game, the experimenter gives an experimental participant a sum of money to distribute—let’s say they give Alice \$100 and ask her to propose giving a portion of this money to a second experimental participant: Bob. Bob is instructed to respond by either accepting Alice’s offer or rejecting it. If Bob rejects Alice’s offer then neither of them get anything. Standard economics predicts that Alice will be self-interested when she plays this game and will aim to offer Bob the very lowest offer that she thinks she will get away with. In this case she would offer Bob \$1, and if Bob is similarly rational he would accept \$1 because \$1 is better than \$0. In reality, however, in a very extensive range of studies of the Ultimatum Game—including studies across cultures, socioeconomic characteristics, and even experiments with monkeys playing the game for juice and fruit—the proposers are remarkably generous in offering much more than the equivalent of \$1. On the other hand, the responders will often reject even relatively generous offers. What is going on? Behavioral economists explain these findings, and other findings from similar games, in terms of our social preferences. We do not like seeing unequal outcomes—we experience inequity aversion, and we experience it in two forms: disadvantageous inequity aversion, and advantageous inequity aversion. Disadvantageous inequity aversion is when we do not want to suffer inequity ourselves. In the Ultimatum Game, Bob will suffer disadvantageous inequity



Queen Elizabeth II of the United Kingdom and Prince Philip, Duke of Edinburgh, during a visit to the remodeled King Edward Court Shopping Centre in Windsor, England, February 2008







aversion when Alice makes a mean offer—and this may lead him to reject offers of relatively large amounts. On the other hand, advantageous inequity aversion is about not wanting to see others around us treated unfairly—so Alice will not make the minimum possible offer to Bob because she reasons that that would be unfair. Unsurprisingly, we worry much more about disadvantageous inequity aversion than advantageous inequity aversion, but both have been demonstrated—across a large number of experimental studies—to have a strong influence on our tendencies toward generosity.

## Social Influences

Linking to these insights around social preferences, behavioral economists have explored some other ways in which social influences affect our decisions and choices. Broadly speaking, these social influences can be divided into informational influences and normative influences (Baddeley, 2018a). Informational influences are about how we learn from others. In situations where we do not know much or are facing a complex and uncertain series of potential outcomes, it makes sense for us to look at what others are doing, inferring that they may know better than we do about the best course of action. Economists have analyzed this phenomenon in terms of updating our estimates of probabilities—and a classic example outlined by Abhijit Banerjee is restaurant choice (Banerjee, 1992). We are new to a city—perhaps visiting as tourists—and we see two restaurants, both of which look similar but we have no way of knowing which is better. We see that one is crowded and the other is empty and—perhaps counter-intuitively—we do not pick the empty restaurant, which might be more comfortable and quieter. Instead, we pick the crowded restaurant. Why? Because we infer that all those people who have chosen the crowded restaurant ahead of the empty restaurant know what they are doing, and we follow their lead—using their actions (the restaurant they choose) as a piece of social information. We respond to these informational influences in a rational way—perhaps not the extreme form of rationality that forms the cornerstone of a lot of economics, but nonetheless sensible—the outcome of a logical reasoning process.

Normative social influences are less obviously rational and are about how we respond to pressures from the groups around us. In explaining these social pressures, behavioral economics draws on key insights from social psychologists, such as Stanley Milgram and Solomon Asch, and their colleagues. Stanley Milgram created controversy with his electric shock experiments. Experimental participants were instructed by an experimenter to inflict what they thought were severe electric shocks on other people hidden from view. The participants in Milgram's experiments could still hear the people who were supposedly receiving the shocks. In fact, these people were just actors but the experimental participants did not know this and a significant number of the participants (not all) were prepared to inflict what they were told were life-threatening levels of shock: the actors pretended to experience severe pain, screaming and at worst in some cases going worryingly quiet after the shocks. Milgram explained the fact that his participants were prepared to act in these apparently ruthless ways as evidence that we are susceptible to obedience to authority. We are inclined to do what we are told, especially when we confront physically and psychologically challenging scenarios. Milgram's evidence was used in part to explain some of the atrocities associated with the Holocaust—in an attempt to answer the puzzle of why so many otherwise ordinary civilians not only observe but also actively engage in atrocities.

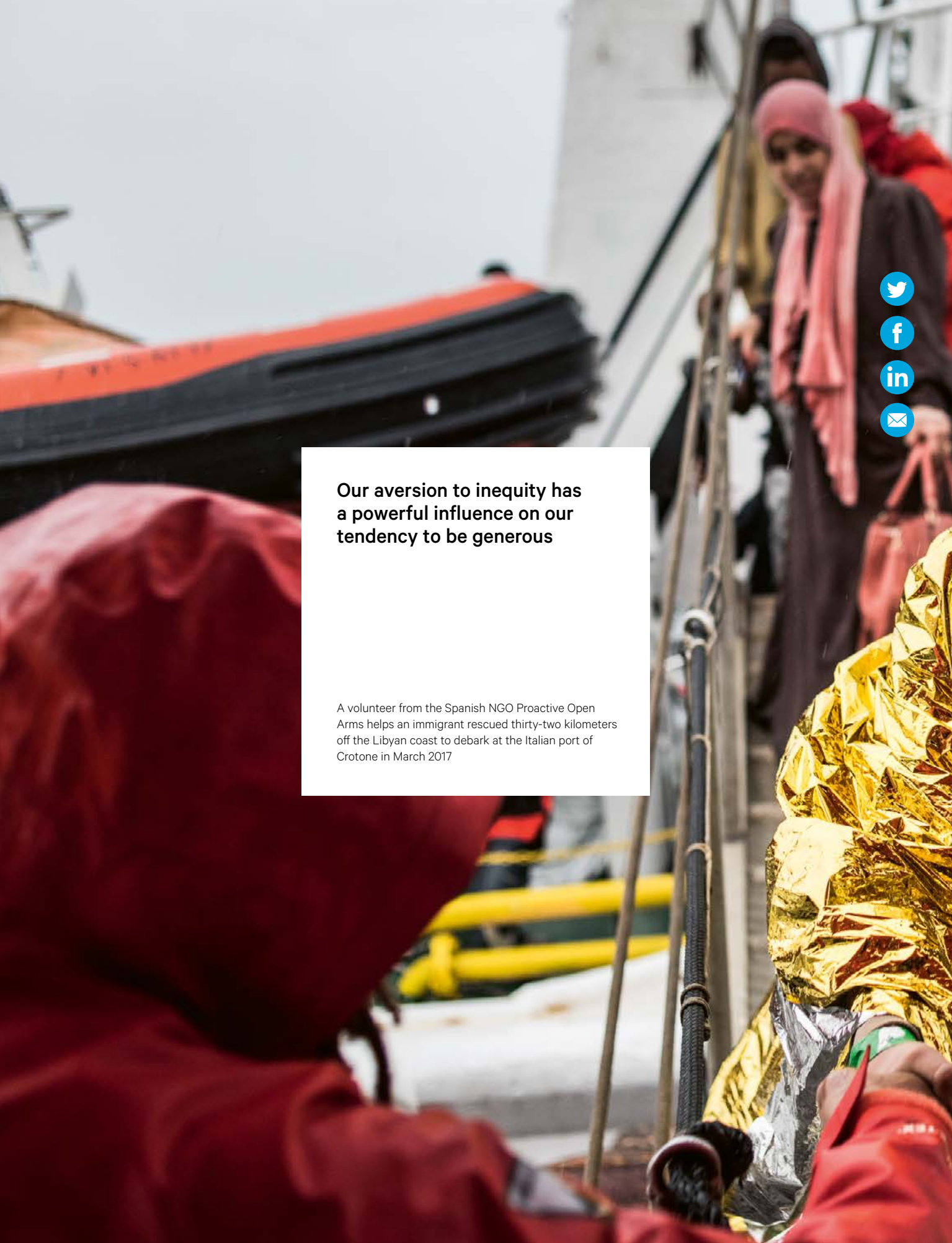
Another influential set of social psychology experiments that have informed behavioral economists include Solomon Asch's experiments (Asch, 1955). He devised a line experiment



to test for conformity: experimental participants were asked to look at a picture of a line and then match it with another line of the same length. This was an easy task, but Asch and his colleagues added a complication by exposing their participants to other people's guesses. Unbeknownst to their participants, the groups deciding about the line lengths in fact included a large number of experimental confederates instructed to lie about the length of the lines. To illustrate with a simple example: imagine that twenty participants are gathered together to complete the line task but nineteen are in cahoots with the experimenter and there is only one genuine participant. If the others all came up with a stupid, wrong answer to this simple question about lines, what would the twentieth, genuine participant do? Asch and his colleagues found that many of the genuine participants (though, tellingly, not all) changed their minds away from the correct answer to give an obviously wrong answer when they saw others making a wrong guess. In other words, many participants seemed inclined to ensure that their answers conformed with the answers from the other participants in their group, without considering that these participants might be mistaken, or lying. The emotional responses of the participants were variable. Those who stuck with their original answers did so confidently. The conformists who changed their answers to fit with the group varied—some experiencing distressing self-doubt; others blaming other participants for their mistakes. Why would a person change their mind to what otherwise would seem like an obviously wrong answer? This experiment does not resolve the rational versus irrational question. It may seem irrational to give the wrong answer just because you see others getting it wrong. The Nobel Prize-winning economist Robert Shiller came up with another explanation, consistent with rational decision-making: perhaps the real participants were thinking that it is much more likely that their single decision was wrong than that nineteen others were wrong. They were balancing the probabilities and coming to the conclusion that the chances that such a large number of other people could be wrong were small and so it made sense to follow them (Shiller, 1995).

## **Social influences can be divided into informational and normative influences. The former are about how we learn from others, while the latter are about how we respond to pressures from groups around us**

More generally, many of us use others' choices and actions to guide our own choices and actions—such as in the restaurant example above. When we copy other people we are using a rule of thumb—a simple decision-making tool that helps us to navigate complex situations, especially situations characterized by information overload and choice overload. In today's world, the ubiquity of online information and reviews is another way in which we use information about others' choices and actions as a guide. For example, when we are buying a new computer or booking a hotel, we will find out what others have done and what others think before deciding for ourselves. In these situations, when thinking through lots of information and many choices is a cognitive challenge, it makes sense to follow others and adopt what behavioral economists would call a herding "heuristic." Following the herd is a quick way to decide what to do. This brings us to the large and influential literature on heuristics and bias, developing out of Daniel Kahneman's and his colleague Amos Tversky's extensive experimental work in this field.



**Our aversion to inequity has  
a powerful influence on our  
tendency to be generous**

A volunteer from the Spanish NGO Proactive Open Arms helps an immigrant rescued thirty-two kilometers off the Libyan coast to debark at the Italian port of Crotona in March 2017











What are heuristics? Heuristics are the quick decision-making rules we use to simplify our everyday decision-making and they often work well, but sometimes they create biases in our decision-making. In other words, in some situations when we use heuristics they lead us into systematic mistakes. The psychologist Gerd Gigerenzer makes the important observation, however, that heuristics are often good guides to decision-making because they are fast and frugal. They often work well, especially if people are given simple techniques to enable them to use heuristics more effectively (Gigerenzer, 2014).

### **When thinking through lots of information and many choices is a cognitive challenge, it makes sense to follow others and adopt what behavioral economists would call a herding “heuristic”**

If you Google behavioral bias today, you will get a long and unstructured list, and, in devising a taxonomy of heuristics and their associated biases, a good place to start is Daniel Kahneman and Amos Tversky’s taxonomy of heuristics—as outlined in their 1974 *Science* paper (Tversky and Kahneman, 1974) and summarized for a lay audience in Kahneman (2011). Kahneman and Tversky identified three categories of heuristic, based on evidence from an extensive range of experiments they had conducted, including the availability, representativeness, and anchoring and adjustment heuristics.

The availability heuristic is about using information that we can readily access—either recent events, first moments, or emotionally vivid or engaging events. Our memories of these types of highly salient information distort our perceptions of risk. A classic example is the impact that vivid and sensationalist news stories have on our choices, linking to a specific type of availability heuristic—the affect heuristic. For example, vivid accounts of terrible plane and train crashes stick in our memory leading us to avoid planes and trains when, objectively, we are far more likely to be run over by a car when crossing the road, something we do every day without thinking too hard about it. We misjudge the risk—thinking plane and train crashes are more likely than pedestrian accidents—and this is because information about plane crashes is far more available, readily accessible, and memorable for us.

The representativeness heuristic is about judgments by analogy—we judge the likelihood of different outcomes according to their similarity to things we know about already. In some of their experiments, Kahneman and Tversky asked their participants to read a person’s profile and judge the likelihood that this profile described a lawyer versus an engineer. They discovered that many of their participants judged the likelihood that a person described was a lawyer or an engineer according to how similar the profile was to their preconceptions and stereotypes about the characteristic traits of lawyers versus engineers.

Anchoring and adjustment is about how we make our decisions relative to a reference point. For example, when participants in Kahneman and Tversky’s experiments were asked to guess the number of African nations in the United Nations, their guesses could be manipulated by asking them first to spin a wheel to give them a number. Those who spun a lower number on the wheel also guessed a smaller number of African countries in the UN.

Another seminal contribution from Kahneman and Tversky emerges from their analyses of heuristics and bias: their own unique behavioral theory of risk—what they call “prospect theory” (Kahneman and Tversky, 1979). They devised prospect theory on the basis of a se-



ries of behavioral experiments which suggested some fundamental flaws in expected utility theory—economists’ standard theory of risk. The differences between these two different approaches to understanding risky decision-making are complex but one of the fundamental features of expected utility theory is that it assumes that people’s risk preferences are stable: if someone is a risk-taker then they are a risk-taker. They will not shift their decisions if the risky choices they are offered are framed in a different way. This connects with three fundamental features of prospect theory: in prospect theory, risk preferences are shifting. People’s risk preferences do shift in prospect theory. For example, they are more inclined to take risks to avoid losses—linking to a key insight from prospect theory: “loss aversion.” Standard economics predicts that whether we are facing losses or gains, we decide in the same way according to the absolute magnitude of the impact for us. In prospect theory, however, people confront losses and gains differently—we worry much more about losses than we do about gains, and one facet of this is that we will take bigger risks to avoid losses than we will to accrue gains. This also links to another key feature of prospect theory. We make decisions according to our reference point—and most often this is the status quo, our starting points. This feature connects directly to the anchoring and adjustment heuristic, which we explored above.

### Time and Planning

A whole other swathe of behavioral economics literature taps into some important insights about our ability to plan our choices and decisions over time. Standard economics predicts that we form stable preferences about time, as we do for risk. This means that it does not matter what time horizon we are considering. If we are impatient, we are impatient, no matter what the context. Behavioral economists overturn this understanding of how we plan and make decisions over time, building on the substantial evidence from psychological experiments that we are disproportionately impatient in the short term—we suffer from what behavioral economists call present bias. We overweight benefits and costs that come sooner relative to those that come later. For example, if we are choosing between spending on our credit card today or tomorrow and comparing this choice with spending on our credit card in a year or a year and a day, then standard economics predicts that our choices should be time consistent: if we prefer to spend today then we should prefer to spend in a year; and if we prefer to spend in a year and a day, then we should also prefer to spend tomorrow. But behavioral experiments show that we are disproportionately impatient in the short term relatively to the longer term: we prefer to spend today over tomorrow, but when planning for the future we prefer to spend in a year and a day than in a year. We overweight immediate rewards. Behavioral economists such as David Laibson have captured this within alternative theories of discounting to that incorporated in standard economics—specifically in the form of hyperbolic discounting (Laibson, 1997). This is more than an academic curiosity because it has significant implications in our everyday lives—in explaining everything from procrastination to addiction. Present bias can explain why we delay actions that are costly or unpleasant. It can also explain a range of bad habits, or lack of good habits. A telling experiment was one conducted by economists Stefano DellaVigna and Ulrike Malmendier in their study of gym-going habits. Looking at a dataset from a real-world gym, they found that some people signed-up for annual contracts and then attended the gym only a handful of times—even though they had been offered pay-as-you-go membership as an alternative (DellaVigna and Malmendier, 2006). Over the course of a year and sometimes longer, these very occasional gym-goers were effectively paying enormous sums per visit when they would not have had



to if they had more accurately forecasted their future behavior when they signed-up for the gym. This is difficult to explain in terms of standard economic analysis, but once behavioral economists allow for present bias, this behavior becomes explicable. Gym-goers plan at the outset to go to the gym many times, but they change their plans when confronted with the immediate choice between going to the gym versus another (more) enjoyable activity.

## **Behavioral experiments show that we are disproportionately impatient in the short term relatively to the longer term: we prefer to spend today over tomorrow, but when planning for the future we prefer to spend in a year and a day than in a year**

Present bias can also explain why we overeat and struggle so hard to give up nicotine, alcohol, and other drugs. There are nuances too—in the way some of us deal with our tendency toward present bias. More sophisticated decision-makers realize that they suffer present bias so they bind their future selves, using what behavioral economists call commitment devices. For example, they might freeze their credit card in a block of ice to stop their future self being tempted into impulsive spending splurges. New businesses have developed around these commitment devices—including online self-tracking tools such as Beeminder. When you sign-up for Beeminder, you set out your goals, and if you fail to meet those goals, Beeminder charges you for your transgression.

A key feature of present bias, and other biases, is that we are not all equally susceptible. Some of us are better at self-control than others and there is a large and growing literature on individual differences, including personality traits, and the role these play in explaining our different susceptibilities to behavioral bias. Learning lessons from psychologists, behavioral economists are using personality tests to help to explain some differences in susceptibility to biases such as present bias. One set of tests now widely used by behavioral economists is the Big Five OCEAN tests—where OCEAN stands for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In their analyses of the impact of individual differences on economic success, Lex Borghans and Nobel Prize-winner James Heckman found, for example, that conscientiousness comes out as being a key trait strongly correlated with success in life, confirming earlier findings from psychologist Walter Mischel—famous for the Marshmallow Experiment—which studied children’s capacity to resist temptation when choosing between one marshmallow today versus two tomorrow: those children who were better able to exert self-control in resisting temptation were also more likely to succeed later in life (Borghans et al., 2008; Mischel, 2014).

### **The Future: Nudging and Beyond**

All these insights from behavioral economics are now changing mainstream economics, and also having a strong impact on policy-making via nudging, as highlighted in the introduction. So, are there new horizons for behavioral economics, or do we know all we need to know? For nudging, more evidence is needed to capture how robust and scalable nudging policies really are—and there has been progress in this direction. Another key area that has been largely neglected until recently is behavioral macroeconomics. British economist John Maynard Keynes pioneered the analysis of psychological influences, particularly social conventions, in financial markets and the implications for macroeconomics more generally—see for example





Pedestrians on a crosswalk are reflected in the facade of a mall in Tokyo's Omotesando shopping district, March 2013





Keynes (1936). Some of Keynes's insights are being reimagined today, for instance by Nobel Prize-winning economists including George Akerlof and Robert Shiller (see Akerlof, 2002; and Akerlof and Shiller, 2009). These insights were complemented by American economist George Katona's macroeconomic insights, especially his analyses of consumer sentiment (Katona, 1975). Katona's influence endures through the University of Michigan's Consumer Sentiment Index—outputs from which are also still being widely used today (see for example Curtin, 2018). A significant hurdle for behavioral macroeconomics, however, is that it is difficult coherently to aggregate into a macroeconomic model the complexities of behavior identified by behavioral economists within a microeconomic context. New methodologies are coming on board however, for example in the form of agent-based modeling and machine learning. If these new methods can be applied successfully in developing coherent behavioral macroeconomic models, then behavioral economics will generate an even more exciting and innovative range of insights in the forthcoming decade than it has in the last.



## Select Bibliography

- Akerlof, George. 2002. "Behavioural macroeconomics and macroeconomic behavior." *American Economic Review* 92(3): 411–433.
- Akerlof, George, and Shiller, Robert. 2009. *Animal Spirits: How Human Psychology Drives the Economy and Why It Matters for Global Capitalism*. Princeton: Princeton University Press.
- Ariely, Dan. 2008. *Predictably Irrational – The Hidden Forces that Shape Our Decisions*. New York: Harper Collins.
- Asch, Solomon. 1955. "Opinions and social pressure." *Scientific American* 193(5): 31–35.
- Baddeley, Michelle. 2017. *Behavioural Economics: A Very Short Introduction*. Oxford: Oxford University Press.
- Baddeley, Michelle. 2018a. *Copycats and Contrarians: Why We Follow Others... and When We Don't*. London/New Haven: Yale University Press.
- Baddeley, Michelle. 2018b. *Behavioural Economics and Finance* (2nd edition). Abingdon: Routledge.
- Banerjee, Abhijit. 1992. "A simple model of herd behavior." *Quarterly Journal of Economics* 107(3): 797–817.
- Borghans, Lex, Duckworth, Angela Lee, Heckman, James J., and Ter Well, Bas. 2008. "The economics and psychology of personality traits." *Journal of Human Resources* 43(4): 972–1059.
- Curtin, Richard. 2018. *Consumer Expectations: Micro Foundations and Macro Impact*. New York/Cambridge: Cambridge University Press.
- DellaVigna, Stefano, and Malmendier, Ulrike. 2006. "Paying not to go to the gym." *American Economic Review* 96(3): 694–719.
- Gigerenzer, Gerd. 2014. *Risk Savvy: How to Make Good Decisions*. London: Penguin Books.
- Gneezy, Uri, and Rustichini, Aldo. 2000. "A fine is a price." *Journal of Legal Studies* 29(1): 1–17.
- Güth, Werner, Schmittberger, Rolf, and Schwarze, Bernd. 1982. "An experimental analysis of ultimatum bargaining." *Journal of Economic Behavior and Organization* 3: 367–388.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Strauss and Giroux.
- Kahneman, Daniel, and Tversky, Amos. 1979. "Prospect theory – an analysis of decision under risk." *Econometrica* 47(2): 263–292.
- Katona, George. 1975. *Psychological Economics*. New York: Elsevier.
- Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. London: Royal Economic Society/Macmillan.
- Laibson, David. 1997. "Golden eggs and hyperbolic discounting." *Quarterly Journal of Economics* 112: 443–478.
- Milgram, Stanley. 1963. "Behavioral study of obedience." *Journal of Abnormal and Social Psychology*. 67: 371–378.
- Mischel, Walter. 2014. *The Marshmallow Test: Why Self-Control Is the Engine of Success*. New York: Little, Brown and Company.
- Shiller, Robert. 1995. "Conversation, information and herd behavior." *American Economic Review* 85(2): 181–185.
- Simon, Herbert. 1955. "A behavioral model of rational choice." *Quarterly Journal of Economics* 69: 99–118.
- Thaler, Richard H. 2016. *Misbehaving – The Making of Behavioural Economics*. London: Allen Lane.
- Thaler, Richard H., and Sunstein, Cass. 2008. *Nudge – Improving Decisions About Health, Wealth and Happiness*. London/New Haven: Yale University Press.
- Tversky, Amos, and Kahneman, Daniel. 1974. "Judgement under uncertainty: Heuristics and bias." *Science* 185: 1124–1121.



**Nancy H. Chau**  
Cornell University

Nancy H. Chau is Professor of Economics in the Charles H. Dyson School of Applied Economics and Management. Professor Chau's research interests fall under three main areas: international trade, regional economics, and economic development. Professor Chau was recently awarded the Alexander von Humboldt Research Fellowship, and the first T. W. Schultz Award of the International Agricultural Economics Association. She is a senior fellow at the Center for Development Research, a research fellow at the Institute for the Study of Labor (IZA Bonn), and member of an expert panel for the Office of the UN High Commissioner for Human Rights. Professor Chau has published widely, in journals such as *Economic Journal*, *International Economic Review*, *Journal of Development Economics*, *Journal of Economic Growth*, *Journal of Economic Theory*, *Journal of Labor Economics*, *Journal of Public Economics*, *Journal of Public Economic Theory* and the *World Bank Economic Review*.



**Ravi Kanbur**  
Cornell University

Ravi Kanbur researches and teaches in development economics, public economics, and economic theory. He is well known for his role in policy analysis and engagement in international development. He has served on the senior staff of the World Bank, including as Chief Economist for Africa. He has also published in the leading economics journals, among which *Journal of Political Economy*, *American Economic Review*, *Review of Economic Studies*, and *Journal of Economic Theory and Economic Journal*. The positions he has held or holds include: President of the Human Development and Capabilities Association, Chair of the Board of United Nations University-World Institute for Development Economics Research, Co-Chair of the Scientific Council of the International Panel on Social Progress, member of the OECD High Level Expert Group on the Measurement of Economic Performance, President of the Society for the Study of Economic Inequality, Member of the High Level Advisory Council of the Climate Justice Dialogue, and Member of the Core Group of the Commission on Global Poverty.

Recommended book: *Contributions to the Economics of International Labor Standards*, Arnab K. Basu and Nancy H. Chau, World Scientific Publishing Co., 2017.

**This overview considers the past, the present, and the future of economic development. It begins with the conceptualization, definition, and measurement of economic development, highlighting that a narrow focus on the economic is inadequate to capture development and even, paradoxically, economic development itself. Key aspects of economic and human development over the past seven decades are then outlined, and the current landscape is described. The paper then considers the future of economic development, highlighting the challenges faced by developing countries, especially the opportunities and risks provided by the recent downward global trend in the share of labor in overall economic activity.**



What is economic development and how has the concept evolved through the years? The economic part of it could be thought to be relatively straightforward. Surely, a steady rise in per capita income as conventionally measured is an anchor, in concept and in reality. It would be odd indeed to describe declining per capita income as economic development. But rising per capita income, while necessary, is certainly not sufficient for development, and even for economic development.

The distribution of this rising income among the population is legitimately in the domain of economic development. Two key features of the distribution of income are inequality and poverty. If average income rises but the inequality of its distribution also increases, then an egalitarian perspective would mark down the latter as a negative aspect of economic development. If poverty, the population below a socially acceptable level of income, also increases then this is another negative mark to be set against rising average income in assessing economic development. Of course, the actual outcome on poverty will depend on an interaction between average income and inequality and which of the two forces dominates empirically.

**If higher average income is accompanied by increasingly unequal distribution, an egalitarian perspective will qualify it as negative. Growing poverty would also contrast negatively with higher average income in any evaluation of economic development**

But identifying economic development purely with income is too narrow a conception. Other aspects of well-being are surely relevant. Education and health outcomes, for example, go beyond income. They are important markers of well-being in their own right, but they influence, and are influenced by, income. High income can deliver an educated and healthy population, but an educated and healthy population also delivers high income. Thus, any assessment of development, and even economic development, needs to take into account a broader range of measures of well-being than simply income and its distribution. Education and health, and their distribution in the population, are important as well.

Distribution is not simply about inequality between individuals. Inequality across broadly defined groups is also a key factor. Gender inequality saps economic development as it suppresses the potential of half the population. Thus, improvements in measures of gender inequality are to be looked for in their own right, but also because of the contributions they make to economic growth and to addressing economic inequality. Similarly, inequalities between ethnic and regional groups stoke social tension and affect the climate for investment and hence economic growth. It is difficult to separate out these seemingly non-economic dimensions from the narrowly economic. Economic development is thus also about development more generally.

A narrow focus on measured market income misses out on use of resources which are not priced appropriately in the market. The most important of these is the environment, especially in the context of greenhouse gas emissions and climate change. Rising national income as conventionally measured does not price in the loss of irreplaceable environmental resources at the national level nor, in the case of climate change, irreversible moves toward catastrophic risks for the planet we live on.





A broader conception of development has been embraced by the international community, first through the Millennium Development Goals (MDGs) of 2000, and then through the Sustainable Development Goals (SDGs) of 2015. The eight MDGs were expanded and modified to seventeen SDGs, which include conventional economic measures such as income growth and income poverty, but also inequality, gender disparities, and environmental degradation (Kanbur, Patel, and Stiglitz, 2018). Indeed, the crystallization and cementing of this broader conceptualization of development, and even of economic development, has been one of the sure advances during the past decade of thinking, and surely represents a move toward a “new enlightenment” in assessing trajectories of achievement. But what have these trajectories been over the past seven decades since World War II? The next section takes up the story.

## The Past<sup>1</sup>

The six decades after the end of World War II, until the crisis of 2008, were a golden age in terms of the narrow measure of economic development, real per capita income (or gross domestic product, GDP). This multiplied by a factor of four for the world as a whole between 1950 and 2008. For comparison, before this period it took a thousand years for world per capita GDP to multiply by a factor of fifteen. Between the year 1000 and 1978, China’s income per capita GDP increased by a factor of two; but it multiplied six-fold in the next thirty years. India’s per capita income increased five-fold since independence in 1947, having increased a mere twenty percent in the previous millennium. Of course, the crisis of 2008 caused a major dent in the long-term trend, but it was just that. Even allowing for the sharp decreases in output as the result of the crisis, postwar economic growth is spectacular compared to what was achieved in the previous thousand years.

**The six decades after the end of World War II, until the crisis of 2008, were a golden age in terms of the narrow measure of economic development, real per capita income. This multiplied by a factor of four for the world as a whole between 1950 and 2008**

But what about the distribution of this income, and in particular the incomes of the poorest? Did they share in the average increase at all? Here the data do not stretch back as far as for average income. In fact, we only have reasonably credible information going back three decades. But, World Bank calculations, using their global poverty line of \$1.90 (in purchasing power parity) per person per day, the fraction of world population in poverty in 2013 was almost a quarter of what it was in 1981—forty-two percent compared to eleven percent. The large countries of the world—China, India, but also Vietnam, Bangladesh, and so on—have contributed to this unprecedented global poverty decline. Indeed, China’s performance in reducing poverty, with hundreds of millions being lifted above the poverty line in three decades, has been called the most spectacular poverty reduction in all of human history.

But the story of the postwar period is not simply one of rising incomes and falling income poverty. Global averages of social indicators have improved dramatically as well. Primary school completion rates have risen from just over seventy percent in 1970 to ninety percent





**By 2013 the percentage of the world's population living in poverty had dropped to one fourth the percentage of 1981: eleven percent compared to the previous forty-two percent**

A fishermen's neighborhood in Mumbai, where the suburbs are changing their appearance thanks to an organization dedicated to improving living conditions for the disadvantaged in India's financial capital. June, 2018





**The international community  
adopted a more global concept  
of development through the  
2000 Millennium Development  
Goals (MDG)**

The opening session of the Millennium Summit at United Nations Headquarters in New York on September 6, 2000. From left to right: the then Secretary General of the UN, Kofi Annan and co-presidents Tarja Halonen (Finland) and Sam Nujoma (Namibia)



now as we approach the end of the second decade of the 2000s. Maternal mortality has halved, from 400 to 200 per 100,000 live births over the last quarter century. Infant mortality is now a quarter of what it was half a century ago (30 compared to 120, per 1,000 live births). These improvements in mortality have contributed to improving life expectancy, up from fifty years in 1960 to seventy years in 2010.

Focus on just income, health, and education hides another major global trend since the war. This has truly been an age of decolonization. Membership of the UN ratcheted up as more and more colonies gained political independence from their colonial masters, rising from around fifty in 1945 to more than 150 three decades later. There has also been a matching steady increase in the number of democracies with decolonization, but there was an added spurt after the fall of the Berlin Wall in 1989, when almost twenty new countries were added to the democratic fold. To these general and well quantified trends we could add others, less easily documented, for example on women's political participation.

With this background of spectacular achievements at the global level, what is to stop us from declaring a victorious past on human progress? The answer is that we cannot, because good global average trends, although they are to be welcomed, can hide alarming counter tendencies. Countries in Africa which are mired in conflict do not have any growth data to speak of, and indeed any economic growth at all. Again in Africa, for countries for which we have data, although the fraction of people in poverty has been falling, the absolute number in poverty has been rising, by almost 100 million in the last quarter century, because of population growth.

A similar tale with two sides confronts us when we look at inequality of income in the world. Inequality as between all individuals in the world can be seen as made up of two components. The first is inequality between average incomes across countries—the gap between rich and poor countries. The second is inequality within each country around its average. Given the fast growth of large poorer countries like India and China relative to the growth of richer countries like the US, Japan, and those in Europe, inequality between countries has declined. Inequality within countries displays a more complex picture, but sharp rises in inequality in the US, Europe, and in China and India means that overall within-country inequality has increased. Combining the two, world inequality has in fact declined overall (Lakner and Milanovic, 2016). The importance of between-nation inequality has fallen from a contribution of four fifths of global inequality a quarter century ago. But its contribution is still not lower than three quarters of total world inequality. These two features, rising within nation inequality in large developing countries, and the still enormous role of between-nation inequality in global inequality, are the other side of the coin from the good news of developing country growth on average in the last three decades.

**Inequality among Earth's inhabitants comprises two elements: the first, which is expressed by each country's average income, reflects the gap between rich and poor countries; the second reflects inequalities within each country in terms of average incomes**

But income growth, if it comes at the expense of the environment, mis-measures improvement in human well-being. Particulate pollution has increased by ten percent over the last





quarter century, with all of its related health implications. The global population under water stress has almost doubled in the last half century, and there has been a steady decline in global forest area over the same period. Global greenhouse gas emissions have increased from under 40 gigatons equivalent to close to 50 gigatons in the last quarter century. On present trends global warming is projected to be around 4°C by 2100, well above the safe level of 1.5°C warming. The consequences of global warming have already begun to appear in terms of an increase in severe weather outcomes.

Thus, the past seven decades have indeed been golden ones for economic development on some measures, and even development more broadly measured. But all is not golden. The trends hide very worrying tendencies which have begun to surface in terms of their consequences, and are shaping the landscape of development we have with us. The next section takes up the story with a focus on the present of economic development.

## The Present

The present of the economic development discourse is, of course, shaped by the trends of the distant and recent past. An interesting and important feature of the current landscape is the shift in the global geography of poverty. Using standard official definitions, forty years ago ninety percent of the world's poor lived in low-income countries. Today, three quarters of the world's poor live in middle-income countries (Kanbur and Sumner, 2012). The fast growth of some large countries, accompanied by rising inequality in these countries, means that the average income increases have not been reflected in poverty reduction to the same extent. So, although these countries have now crossed the middle-income category boundary, which depends on average income, they still have large absolute numbers of poor people. These poor in middle-income countries vie with the poor in poor countries for global concern and attention.

This disconnect between a person being poor and their country being poor is shaking up the global development assistance system, which was built on the notion that the bulk of the world's poor lived in poor countries. This is manifested in the “graduation” criteria used by most aid agencies, whereby aid is sharply reduced and then cut off when a country's average income crosses a threshold, typically related to middle-income status. It raises the question posed by Kanbur and Sumner (2012): “Poor countries or poor people?” The response has been, by and large, to stay with the average income criteria. This has led to and will increasingly lead to a dichotomy between very poor countries, often mired in conflict, and middle-income countries where, in fact, the bulk of the world's poor now live. Thus, if the World Bank's soft loan arm sticks to its graduation criteria, it will in effect disengage from the vast majority of the world's poor, while focusing on the poorest countries in the world. This disengagement is difficult to justify on ethical grounds, but also difficult to understand if middle-income countries are also the source of global environmental problems and, for some of them, the source of conflict-based migration.

Migration, conflict-based and economic, brings us to another important feature of the present landscape of economic development, one which is the result of past trends and which will surely have global implications for the future. Rising inequality in rich countries has intersected with increased migration pressures from poor countries. Despite the closing of the gap between rich and poor countries because of the fast growth of some poor countries, the gap is still enormous, both on average and especially so for the poorest countries who have not grown as fast. These gaps have combined with increased pressures because of armed conflict and exacerbated by environmental stress.



Ben Bernanke, president of the United States Federal Reserve between 2006 and 2014, lecturing undergraduate seniors at Harvard University, Cambridge, Massachusetts, in June 2008





The hollowing out of the middle class in rich countries has coincided with greater immigration, leading to a toxification of democratic politics in these countries and the rise of far-right, nativist, and xenophobic tendencies in the body politic (Kanbur, 2018). The election of Trump, the vote for Brexit, and the entry of Alternative für Deutschland into the German Parliament are only the most obvious outward manifestations of the current malaise of the body politic. Nor is this just an issue in rich countries. The anti-migrant mob violence in South Africa and ethnic conflict in countries such as Myanmar are part of the same pattern of migration tensions which color economic development today.

The current terrain of economic development has clearly been influenced by the great financial crisis of 2008. Most recently, the global crisis has proved disruptive to development gains, although the losses can be said to have been mainly concentrated in the rich countries. But the reactions and the backlash now apparent in rich countries are having and will have consequences for economic development in poor countries. Further, the genesis of the crisis exposed fault lines in the economic model pursued by rich countries, with wholesale deregulation of markets and especially of banking and capital flows.

## **The hollowing out of the middle class in rich countries has coincided with greater immigration, leading to a toxification of democratic politics in these countries and the rise of far-right, nativist, and xenophobic tendencies in the body politic**

The current state of affairs and ongoing debates relate back to the trajectory of thinking since the fall of the Berlin Wall in 1989. It will be recalled that in a famous statement of the time the events were characterized as marking “the end of history” (Fukuyama, 1989), meaning by this that liberal democracy and free markets had won the battle of ideas. But, as noted by Kanbur (2001), “the end of history lasted for such a short time.” The financial crisis of 1997, emanating from the newly liberalized capital markets of East Asia, was a warning shot. The financial crisis of 2008, emanating in the deregulated financial markets of the US and Europe, led to the world global depression since the 1930s.

The world as a whole is only just recovering from this catastrophe. Its effect on economic thinking has been salutary. Queen Elizabeth II of the United Kingdom famously asked British economists why they did not see it coming. The response from Timothy Besley and Peter Hennessy was that: “So in summary, Your Majesty, the failure to foresee the timing, extent and severity of the crisis and to head it off, while it had many causes, was principally a failure of the collective imagination of many bright people, both in this country and internationally, to understand the risks to the system as a whole” (quoted in Kanbur, 2016). But the risks to the system as a whole were magnified by the deregulatory stance of policy makers in the early 2000s, still basking in the “end of history” narrative of the turn of the millennium. It is to be hoped that the lessons of the devastating crisis of 2008 will not be forgotten as we go forward.

Thus the crisis of 2008 sits atop, and sharpens, negative aspects of trends identified in the previous section and shapes the present and future prospects. These future prospects are taken up in the next section.



The past and present of economic development sets the platform for the long-term future. Environmental degradation and climate change will surely worsen development prospects and ratchet up conflict and environmental stress-related migration. The issues here have been well debated in the literature (see for example, Kanbur and Shue, 2018). And the actions needed are relatively clear—the question is rather whether there is the political will to carry them out.

Beyond challenges that arise due to ecological change and environmental degradation, another prominent challenge that has arisen since the 1980s is the global decline in the labor share. The labor share refers to payment to workers as a share of gross national product at the national level, or as a share of total revenue at the firm level. Its downward trend globally is evident using observations from macroeconomic data (Karababounis and Neiman, 2013; Grossman et al., 2017) as well as from firm-level data (Autor et al., 2017). A decline in the labor share is symptomatic of overall economic growth outstripping total labor income. Between the late 1970s and the 2000s the labor share has declined by nearly five percentage points from 54.7% to 49.9% in advanced economies. By 2015, the figure rebounded slightly and stood at 50.9%. In emerging markets, the labor share likewise declined from 39.2% to 37.3% between 1993 and 2015 (IMF, 2017). Failure to coordinate appropriate policy responses in the face of these developments can spell troubling consequences for the future of economic development. Indeed, the decline in labor share despite overall economic progress is often seen as fuel that has fanned the fire of anti-immigration and anti-globalization backlashes in recent years, threatening a retreat of the decades-long progress made on trade and capital market liberalization worldwide.

It should be noted that the labor share and income inequality are inextricably linked. Indeed, the labor share is frequently used as a measure of income inequality itself (for example, Alesina and Rodrik, 1994). Understanding the forces that determine the labor share has been a singularly important aspect of the landscape of economic development. Indeed, this quest has guided trade and development economics research for decades, during which time the forces of globalization and its many nuanced impacts on the labor share have been fleshed out (Bardhan, 2006; Bourguignon, 2017).

Yet, there are good reasons to take the view that canonical economic models often do not offer predictions consistent with the current pattern of labor share decline in the global economy. Notably, behind the veil of global labor share decline is in fact a tremendous amount of underlying diversity in the direction of change of the labor share at the country level, with emerging and advanced economies at both ends of the spectrum (Karababounis and Neiman, 2013). Such observations are contrary to the canonical prediction of economic models based on the assumptions of constant technologies, perfect competition, and no market imperfections. Guided by these assumptions, the standard prediction is that workers in relatively labor abundant countries should strictly benefit from exposure to world trade in both absolute terms and relative to owners of other inputs of production. In stark contrast, however, after taking on the role as the world's largest factory, China has experienced one of the most significant rates of decline in labor share since 1993 (IMF, 2017).

A search for additional forces that may be in play is clearly warranted.<sup>2</sup> To this end, the trajectory of the global labor share sits at the confluence of three major shifts in the defining features of developing and developed economies. These include: (i) the adoption of labor-saving technological change; (ii) the shift in importance of employer market power; and (iii) the growing prevalence of alternative modes of employment in the labor market.

Labor-saving technological change is a key driver in the recent global labor share decline





(IMF, 2017). The reasons for firms and producers to embrace such a change are many, including a reduction in the price of investment goods and informational technology investment (Karababounis and Neiman, 2013), and the advent of robotics in the manufacturing process (Acemoglu and Restrepo, 2018), for example. Already, advanced economies do not have a monopoly over the adoption of labor-saving technological change. Indeed, China has put in place more robots in manufacturing than any other country according to recent estimates (Bloomberg News, 2017). The implication of labor-saving technological change on labor income is not obvious, however, as it juxtaposes the overall productivity gains that arise from the use of labor-saving technical change, with its potential adverse consequences on unemployment. In the end, whether workers benefit from labor-saving technological change will depend on how quickly productivity gains translate into wage gains (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018; Chau and Kanbur, 2018).

## **An important problem arose in the 1980s: the worldwide decline in the workers' payment as a share of gross national product on a national level, or as a share of total revenue at the firm level**

It is here that additional research can potentially reap significant dividends in furthering our understanding of how developing country markets function and how they respond to shocks. Some important mediating factors have already been identified. These include existing labor market distortions that may skew decision-making about technological change (Acemoglu and Restrepo, 2018), and search friction in the labor market and the resulting possibility of complex distributional responses to technological change (Chau and Kanbur, 2018). Further, policy responses to labor-saving technical change need to be developed and implemented, including perhaps public investment in research into developing efficient labor using technology (Atkinson, 2016; Kanbur, 2018).

In addition to national- or market-level differences in the labor share, recent firm-level evidence has inspired a surge in studies showing that employer market power can give rise to systematic differences in the labor share across firms with heterogeneous productivity levels (for example, Melitz and Ottaviano, 2008). It is by now well known that globalization disproportionately favors high-productivity firms. The ascendance of superstar firms in recent years in the US, with their demonstrably higher propensities to adopt labor-saving technologies, provides an excellent example of how industrial organizational changes can impact the overall labor share (Autor et al., 2017). Employer market power has become a fact of life in emerging markets as well (for example, Brandt et al., 2017). In the course of economic development, does the shift in importance of large firms disproportionately favor the adoption of labor-saving technologies (Zhang, 2013)? Or do they, in fact, value worker morale and pay higher wages (Basu, Chau, and Soundararajan, 2018)? These are critical questions that can inform a host of policy issues going forward, from the desirability of minimum wages to facilitate better wage bargains to be struck for workers, to the use of competition policies as a tool for economic development, for example.

Compounding these shifts in technologies and industrial organization, labor market institutions in emerging markets have also seen significant developments. Present-day labor contracts no longer resemble the textbook single employer single worker setting that forms



the basis for many policy prescriptions. Instead, workers often confront wage bargains constrained by fixed-term, or temporary contracts. Alternatively, labor contracts are increasingly mired in the ambiguities created in multi-employer relationships, where workers must answer to their factory supervisors in addition to layers of middleman subcontractors. These developments have created wage inequities within establishments, where fixed-term and subcontracted workers face a significant wage discount relative to regular workers, with little access to non-wage benefits. Strikingly, rising employment opportunities can now generate little or even negative wage gains, as the contractual composition of workers changes with employment growth. The result can be a downward spiral in worker morale (Basu, Chau, and Soundararajan, 2018). These developments suggest that a decline in labor share generated by contractual shifts in the labor market can ultimately have adverse consequences on the pace of overall economic progress. Attempts to address wage inequities between workers within establishments is a nascent research area (Freeman, 2014; Basu, Chau, and Soundararajan, 2018), and what is intriguing here is the possibility that we now have a set of circumstances under which inequality mitigating policies, by raising worker morale, may end up improving overall efficiency as well.

## **The ascendance of superstar firms with a propensity to adopt labor-saving technologies provides an excellent example of how industrial organizational changes can impact labor's overall share of the GNP**

We began this chapter by emphasizing the joint importance of overall economic progress and income inequality as metrics of development. Our brief look at the future of the economic development landscape sheds light on the critical importance of bringing together multiple perspectives in our understanding of how these two metrics of development are codetermined. Doing so opens up new policy tools (for example, competition policies and technology policies), new reasons for (non-)intervention (for example, workers' morale consequences of wage inequities), and, perhaps equally important, new policy settings where equity and efficiency are no longer substitutes for each other.

### **Conclusion**

Looking back over the past seven decades since the end of World War II, economic development presents us with a string of contradictions. There have been unprecedented rises in per capita income, with many large developing countries crossing the threshold from low-income to middle-income status. These income increases have been accompanied by equally unprecedented improvements in income poverty and in education and health indicators.

But at the same time there is palpable anxiety about the development process, its sustainability, and its implications for the global economy. Despite the fast increases in income in poorer countries, gaps between them and rich countries remain large. Together with conflict and environmental stress, this has led to migration pressures, particularly for richer countries but also for better-off developing countries. The combination of migration pressures and



Moments before Apple announces a product at its new headquarters in Cupertino, California, on September 12, 2018, just one year after launching its iPhone X, the most expensive smartphone on the market



rising inequality has led to the toxic rise of illiberal populist politics which is threatening postwar democratic gains.

While environmental and climate change, and rising inequality in general, have been much discussed, we have highlighted a particular source of rising inequality as an ongoing threat to economic development. The falling share of labor in the economy is set to continue and unless counteracted by strong policy measures will threaten inclusive development in the coming decades.

We have also highlighted how thinking in economics has responded to the underlying forces of change. There has been a broadening of the concept of development beyond the narrowly economic. The roots of the great financial crisis of the end of the first decade of the new millennium have also been scrutinized and, hopefully, some lessons have been learned. And attention is turning to understanding the inexorable decline in labor's share. Whether all this adds up to a New Enlightenment in economic thinking is something the next decades of development will reveal.





## Notes

1. This section draws heavily on data and sources referenced in Chapter 1 of Fleurbaey et al. (2018).
2. Our discussion of the literature on the global labor share decline is necessarily selective here. See Grossman et al. (2017) for a notable counterpoint where a global productivity slowdown in the presence of capital and skill complementarity and endogenous human capital accumulation can also give rise to a global decline in the labor share.

## Select Bibliography

- Acemoglu, Daron, and Restrepo, Pascual. 2018. “The race between man and machine: Implications of technology for growth, factor shares and employment.” *American Economic Review* 108(6): 1488–1542.
- Acemoglu, Daron, and Autor, David. 2011. “Skills, tasks and technologies: Implications for employment and earnings.” In *Handbook of Labor Economics*, Orley Ashenfelter and David Card (eds.). Amsterdam: Elsevier-North, 4: 1043–1171.
- Alesina, Alberto, and Rodrik, Dani. 1994. “Distributive politics and economic growth.” *Quarterly Journal of Economics* 109: 465–490.
- Atkinson, Anthony B. 2016. *Inequality: What Can be Done?* Cambridge, MA: Harvard University Press.
- Autor, David, Dorn, David, Katz, Lawrence, Patterson, Christina, and van Reenen, John. 2017. “The fall of the labor share and the rise of superstar firms.” NBER Working Paper No. 23396.
- Bardhan, Pranab. 2006. “Does globalization help or hurt the world’s poor?” *Scientific American* 294(4): 84–91.
- Basu, Arnab K., Chau, Nancy H., and Soundararajan, Vidhya. 2018. “Contract employment as a worker discipline device.” IZA Discussion Paper No. 11579.
- Bloomberg News. 2017. “China’s robot revolution may affect the global economy—Automation impact on wages isn’t showing in the data – yet.” <https://www.bloomberg.com/news/articles/2017-08-22/china-s-robot-revolution-may-weigh-on-global-rebalancing> (accessed on September 19, 2018).
- Brandt, Loren, Van Beisebroeck, Johannes, Wang, Luhang, and Zhang, Yifan. 2017. “WTO accession and performance of Chinese manufacturing firms.” *American Economic Review* 107: 2784–2820.
- Bourguignon, François. 2017. *The Globalization of Inequality*. New Jersey: Princeton University Press.
- Chau, Nancy, and Kanbur, Ravi. “Employer power, labor saving technical change, and inequality.” Cornell University, Dyson Working Paper 2018–04.
- Fleurbaey, Marc, Bouin, Olivier, Salles-Djelic, Marie-Laure, Kanbur, Ravi, Nowotny, Helga, and Reis, Elisa. 2018. *A Manifesto for Social Progress*. Cambridge: Cambridge University Press.
- Freeman, Richard B. 2014. “The subcontracted labor market.” *Perspectives on Work* 18: 38–41.
- Fukuyama, Francis. 1989. “The end of history?” *The National Interest* 16: 3–18. ISSN 0884-9382. JSTOR 24027184 (accessed on September 19, 2018).
- Grossman, Gene M., Helpman, Elhanan, Oberfield, Ezra, and Sampson, Thomas. 2017. “The productivity slowdown and the declining labor share: A neoclassical exploration.” NBER Working Paper 23853.
- IMF. 2017. *World Economic Outlook*. International Monetary Fund. Washington, DC.
- Kanbur, Ravi. 2001. “Economic policy, distribution and poverty: The nature of disagreements.” *World Development* 29(6): 1083–1094.
- Kanbur, Ravi. 2016. “The end of laissez faire, the end of history and the structure of scientific revolutions.” *Challenge* 59(1): 35–46.
- Kanbur, Ravi. 2018. “Inequality and the openness of borders.” Cornell University, Dyson Working Paper 2018–10.
- Kanbur, Ravi, Patel, Ebrahim, and Stiglitz, Joseph. 2018. “Sustainable development goals and measurement of economic and social progress.” Cornell University, Dyson Working Paper 2018–12.
- Kanbur, Ravi, and Shue, Henry (eds.). 2018. *Climate Justice: Integrating Economics and Philosophy*. Oxford: Oxford University Press.
- Kanbur, Ravi, and Sumner, Andy. 2012. “Poor countries or poor people? Development Assistance and the New Geography of Global Poverty.” *Journal of International Development* 24(6): 686–695.
- Karabarbounis, Loukas, and Neiman, Brent. 2013. “The global decline of the labor share.” *Quarterly Journal of Economics* 129: 61–103.
- Lakner, Christoph, and Milanovic, Branko. 2016. “Global income distribution: from the fall of the Berlin Wall to the great recession.” *The World Bank Economic Review* 30(2): pp. 203–232. <https://doi.org/10.1093/wber/lhv039> (accessed on September 19, 2018).
- Melitz, Marc J., and Ottaviano, Gianmarco I. P. 2008. “Market size, trade and productivity.” *Review of Economic Studies* 75: 295–316.
- Zhang, Hongsong. 2013. “Biased technology and contribution of technological change to economic growth: Firm-level evidence.” 2013 Annual Meeting, August 4–6, 2013, Washington, DC. 150225, Agricultural and Applied Economics Association.



**Vivien A. Schmidt**  
Boston University

Vivien A. Schmidt is Jean Monnet Professor of European Integration and Professor of International Relations and Political Science in the Pardee School at Boston University, where, until 2017, she served as the Founding Director of its Center for the Study of Europe. Her latest honors and awards include being named Chevalier in the Legion of Honor and receiving a Guggenheim Foundation Fellowship for a transatlantic investigation of the populist “rhetoric of discontent.” She has held visiting professorships and fellowships at a wide range of European institutions, including LUISS University in Rome, the Free University of Brussels, the Copenhagen Business School, the Free University of Berlin, Sciences Po Paris, the European University Institute in Florence, and Oxford University. Her recent books include the forthcoming *Resilient Liberalism in Europe’s Political Economy* (co-edited, 2013), *Democracy in Europe* (2006)—named by the European Parliament as one of the “100 Books on Europe to Remember”—and *The Futures of European Capitalism* (2002).

Recommended book: *How Democracies Die*, Steven Levitsky and Daniel Ziblatt, Crown, 2016.

**Not everything in this “transcendent decade” is taking us toward a New Enlightenment. Governance and democracy face particular challenges from the election of populist leaders. The voices of dissent may speak in different languages, but they convey the same sets of messages, draw from the same range of sources, and articulate their outrage in similar ways, using rhetorical strategies that reject experts, excoriate the media, and demonize conventional policies, politicians, and parties. These dissenting voices, long isolated on the margins, now constitute an existential threat to the long-standing consensus on how to conduct politics and manage the economy in liberal democracies. They challenge the institutional commitments of political liberalism to tolerant, balanced governance and the ideational preferences of economic neoliberalism for open borders and free trade.**



Not everything in this “transcendent decade” is taking us toward a New Enlightenment. Governance and democracy face particular challenges. The rise of what is often called “populism” constitutes the biggest challenge to political stability and democracy seen since the 1920s or 1930s (Müller, 2016; Judis 2016; Levitsky and Ziblatt, 2018; Eichengreen, 2018).

The British vote to exit the EU followed by the election of Donald Trump to the US presidency took mainstream politics (and pundits) by surprise. And this was only the beginning of the tsunami that has since swept across continental Europe. Emmanuel Macron’s victory in the French presidential election turned out to be only a momentary reprieve, as the populist extremes became (part of) governing majorities in central and eastern Europe, Austria, and Italy, while gaining ground everywhere else. In some countries, most notably Hungary and Poland, populist governments are undermining the basic institutions of liberal democracy. And in so doing, they seek to emulate the anti-democratic, authoritarian drift of their neighbors to the East, including Turkey and Russia.

The voices of populist dissent may speak in different languages but they convey the same sets of messages: against immigration and open borders, globalization and free trade, Europeanization and the euro. They draw from the same range of sources: the economics of those feeling “left behind,” the sociology of those worried about the “changing faces of the nation,” or the politics of those who want to “take back control.” Most also articulate their outrage in similar ways, using rhetorical strategies that deploy “uncivil” words and “fake news” to create “post-truth” environments that reject experts, excoriate established media, and demonize conventional political elites and parties. These dissenting voices, long isolated on the margins, now constitute an existential threat to the long-standing consensus on how to conduct politics and manage the economy in liberal democracies. They challenge the institutional commitments of political liberalism to tolerant, balanced governance and the ideational preferences of economic neoliberalism for open borders and free trade.

In short, over the past decade, what had long looked like disparate groups of dissatisfied citizens marginalized on the sidelines of mainstream politics, supporting motley crews of anti-establishment leaders and small extremist parties, has coalesced into an all-out assault on liberal democracy and democratic capitalism. The main question to answer therefore is: why and how have populists succeeded today in channeling public fear and anger in ways that have gained them unparalleled influence and even propelled some of their anti-system parties into power?

## **Not everything in this “transcendent decade” is taking us toward a New Enlightenment. Governance and democracy face particular challenges. The rise of populism constitutes the biggest challenge to political stability and democracy seen since the 1920s or 1930s**

Potential answers abound. For some, “it’s the economy, stupid,” especially following the 2008 US financial crisis and the 2010 EU sovereign debt crisis. For others, it is the “cultural backlash” of citizens clinging to their social status and worried about rising immigration. For yet others, it follows from the hollowing out of mainstream political institutions and party politics, accompanied by the political frustration of citizens who feel their voices are not heard and their complaints ignored by national politicians and supranational technocrats. So which is right?



All, in fact, offer valuable insights into the many different reasons for the populist tsunami. But although these analyses help us understand the sources of citizens' underlying anger, they cannot explain why populism has surfaced today with such intensity and in so many different forms in different national contexts. For an answer to why now, in these ways, with this kind of populism, we need to delve more deeply into the nature and scope of populism. This means taking seriously the substantive content of populist leaders' ideas and discourses championing "the people" against the elites while contesting institutionalized expertise. It requires investigating populists' discursive processes of interaction, such as their strategies of communication using new media to consolidate activist social movements and party networks as well as traditional media to disseminate their messages more widely. But any explanation of populist success also demands consideration of the electoral promises generally long on anti-system complaints but vague on policies (at least when outside power); investigating how populist electioneering may affect mainstream politics; and, of course, examining what happens if and when populists gain power.

This article begins with a discussion of the sources of populist discontent, economic, socio-cultural, and political, along with the precipitating role of recent crises. It follows with an analysis of the defining features of today's populism, and how these have constituted existential challenges to liberal democracy. These include the style of populist leaders' discourse, the post-truth content and processes of populist communication, and the connections between populists' promises and their actions. The conclusion asks whether this is a momentary phenomenon, or a new moment in history, and asks what forces may determine the future possibilities.

## The Sources of Populism

How do we explain the meteoric rise of populism over the past decade? For this, we need first to consider the sources of discontent. These are economic, resulting from rising inequalities and socioeconomic deprivations since the 1980s; sociological, related to concerns about status, identity, and nationhood in a context of increasing levels of immigration; and political, generated by citizens' growing dissatisfaction with mainstream politics and policies and loss of trust in government and political elites.

**Economic Sources of Discontent** The economic sources of populism are wide-ranging. They include the rise of inequality due to the accumulation of capital by the "one percent," famously investigated by Thomas Piketty (2014), accompanied by an increase in poverty due to regressive taxation plans and cost-cutting that transformed the postwar welfare state into a less generous system, with lower pensions and less security (Hacker, 2006; Hemerijck, 2013). Moreover, globalization has created a wide range of "losers" in de-industrialized areas, while generating a sense of insecurity for the middle classes, worried about losing jobs and status (Prosser, 2016), or joining the "precariat" (Standing, 2011). The economic disruptions from globalization, in particular the shift of manufacturing from advanced to developing countries, have led to more and more people feeling "left behind" (Hay and Wincott, 2012), and produced a "race to the bottom" of lower skilled groups, especially of younger males (Eberstadt, 2016).

Underpinning these socioeconomic problems is the resilience of neoliberal ideas (Schmidt and Thatcher, 2013). These began by promoting global free trade and market liberalization in the 1980s and ended with the triumph of financial capitalism and "hyper-globalization"





(Stiglitz, 2002; Rodrik, 2011). The financial crisis that began in 2007/08 did little to moderate such ideas, while in the euro crisis, the EU's "ordo-liberal" ideas promoting austerity policies have had particularly deleterious consequences, including low growth, high unemployment (in particular in southern Europe), and rising poverty and inequality (Scharpf, 2014).

The economic sources of populist discontent are many, then. But they leave open a number of questions. For example, why did populism rise in eastern Europe despite an unprecedented economic boom powered by globalization and EU integration? Why in Sweden did populists not emerge after the drastic 1992 crisis but over the course of one of Europe's most remarkable recoveries? And Italy has seen worse economic crises before, so why now? Finally, the new "losers" of globalization have been angry about their loss of income and status ever since the triumph of neoliberalism since the 1980s, so why has their unhappiness translated itself into this set of political attitudes and/or political action today? Why, holding these views, did the challenge of populist parties not come sooner?

**Socio-Cultural Sources of Discontent** Explanations for the rise of populism are not just economic; they are also socio-cultural. The populist backlash has been fueled by another aspect of neoliberal globalization: cross-border mobility and the increases in immigration. Nostalgia for a lost past along with fear of the "other" has resulted in the targeting of immigrant groups (Hochschild and Mollenkott, 2009). Certain groups feel their national identity or sovereignty to be under siege in the face of increasing flows of immigrants (Berezin, 2009; McClaren, 2012). And this is often accompanied by rising nativist resentments tied to perceptions that "others"—immigrants, non-whites, women—are "cutting in the line," and taking the social welfare benefits they alone deserve (Hochschild, 2016). Welfare "patriotism" or "chauvinism" has been rearing its head not only on the right side of the spectrum in the US, the UK, or in France but also on the left, in Nordic countries, notably in Denmark.

## **The populist voices of dissent may speak in different languages, but they convey the same sets of messages: against immigration and open borders, globalization and trade, Europeanization and the euro**

Discontent over immigration may undoubtedly also stem from the socioeconomic problems of those "left behind," worried about loss of jobs to immigrants, and unwilling to reward them with welfare benefits. But the socioeconomic can easily be conjoined with the socio-cultural, as worries about loss of jobs combine with fears of loss of status (Gidron and Hall, 2017). These are the people—older, less educated, white, male—whose worldview is threatened by changing demographics resulting from rising immigrant populations. Often, these are the very same people who are equally troubled by intergenerational shifts to post-materialist values such as cosmopolitanism and multiculturalism (Inglehart and Norris, 2016). They can be people who are well off financially, but subscribe to socially conservative philosophies and/or oppose socially liberal policy programs. These are the people who, while they may remain in favor of economic liberalism, focused on ideas about individual responsibility in the economic realm, reject social liberalism.

Social liberalism underpins ideas about individuals' rights to self-determination, which include the expectation of respect for differences not only involving race and ethnicity but



Thousands of emigrants are escorted by the police as they march along the Slovenian-Croatian border on the way to a transit camp in the Slovenian town of Dobova, in October 2015. That year, Europe suffered the greatest refugee crisis since World War II





also gender, and which have been accompanied by expectations of “political correctness” in language. Particularly contentious have been questions of women’s rights when related to abortion and LGBT rights when involving gay marriage and child adoption. Such questions have played themselves out in the US in particular, including the “bathroom” wars in high schools (about which bathrooms transsexuals and non-gender-identifying may use). Such “identity politics” of the left has sometimes been blamed for right-wing conservatives’ openness to populism on the extreme right (for example, Lilla, 2017).

The various socio-cultural counter-politics of identity provide another plausible explanation for the rise of populism. But here, too, the question of “why now?” remains. This kind of politics has been around for a very long time, fed by ethnocentric definitions of “us” versus “the other,” most notably theorized by Carl Schmitt. After all, particular fears and negative perceptions related to immigration have been around for decades, and more recently at least since the advent of demographic decline, the rise of terrorism, and the mass migration of millions of poor east Europeans (including almost a million Muslims from Bosnia and Albania). And further, why is the socio-cultural demand for populism so acute in some countries affected by mass migration (for example, Germany, Sweden, Denmark, France) but not in others (such as Spain)?

**Political Sources of Discontent** Finally, the discontents are also political, as people feel their voices no longer matter in the political process. In some cases, citizens feel that they have lost control as a result of globalization and/or Europeanization—that powerful people at a distance make decisions that have effects over their everyday lives that they do not like or even understand (Schmidt, 2006, 2017). These include not just global or regional decision-making bodies but also big businesses able to use the political system to their advantage, whether in not paying taxes (for example, Apple) or to get the regulations they want, regardless of their effects on social and environmental policies (Hacker and Pierson, 2010).

Popular disaffection is also locally generated, related to national political systems. Some issues are related to policies. Political parties have increasingly appeared to be unresponsive to their constituencies’ concerns, delivering policies that are seen to serve the elites rather than the ordinary citizen (Berman, 2018). Others stem from structural changes in political institutions. Citizens’ ability to express their disenchantment has, ironically, been amplified by the “democratization” of the electoral rules. By opening up access through primaries and referenda, where the most dissatisfied tend to be more motivated to turn out to vote, party leadership contests have largely brought victory for representatives of more extreme positions. This has in turn weakened political parties as representative institutions at the same time that it has made it more difficult to forge alliances “across the aisle” (Rosenbluth and Shapiro, 2018). Additionally, the supranationalization of decision-making in global and/or European institutions has also had its toll on mainstream party politics, by hollowing it. Political leaders find themselves with the predicament of being forced to choose between being responsive to citizens, as their elected representatives, or being responsible by honoring supranational commitments (Mair, 2013).

Politics pure and simple also matters, of course. Mainstream political parties have seemed at a loss with regard to how to respond to populist challengers on the right and on the left. The center-right’s political strategy has until relatively recently entailed a refusal to govern with the extreme right at the same time that it has often taken up their issues in attempts to reclaim their constituencies, in particular with regard to immigration. And while the center right has thus appeared to chase after the extreme right on the hot-button issues, the center left has frequently seemed to chase after the center right on those self-same issues.



Complicating matters for the European Union is the supranational nature of decision-making, and how this has affected national politics. A major shift in the structure of national politics across Europe has occurred as a result of new electoral divides. These involve crosscutting cleavages between traditional political divisions based on adherence to right/left political parties and newer identity-related divisions based on more closed, xenophobic, and authoritarian values versus more open, cosmopolitan, and liberal values (Hooghe and Marks, 2009; Kriesi et al., 2012). The issues most in focus for the xenophobic/authoritarian side of the division began with immigration. But increasingly over the years, the European Union has become an equally politicized issue, as citizens have gone from the “permissive consensus” of the past to the current “constraining dissensus” (Hooghe and Marks, 2009). Public opinion surveys and polls clearly chart citizens’ loss of trust in political elites and of faith in their national democracies, let alone the EU (Pew and Eurobarometer polls, 2008–18).

**Citizens’ ability to express their disenchantment has, ironically, been amplified by the “democratization” of the electoral rules. Primaries and referenda, where the most dissatisfied tend to be more motivated to turn out to vote, have largely brought victory for representatives of more extreme positions**

In the European Union, multilevel governance puts great strain on member-state democracies, albeit each for different reasons of history, culture, and politics (Schmidt, 2006). Note, however, that the citizens’ feelings of disenfranchisement (and the realities) are not only due to the EU’s multilevel political system. While Brexit was probably the *sumum* of the EU’s populist revolt (until the Italian election of March 2018, when Eurosceptic parties gained a governing majority), Trump’s election in the US was fueled by very much the same sentiments. They are in large part a consequence of the growing supranationalization of decision-making in an era of globalization, where governments have exchanged national autonomy for shared supranational authority in order to regain control over the forces they themselves unleashed through national policies of liberalization and deregulation (see, for example, Schmidt, 2002; de Wilde and Zürn, 2012). And with liberalization and deregulation, fueled by neoliberal philosophies (Schmidt and Thatcher, 2013), also came technocratic decision-making, which promoted the depoliticization of policies and processes, along with the downgrading of politics (De Wilde and Zürn, 2012; Fawcett and Marsh, 2014). As a result, mainstream politics has found itself under attack from two sides: the rise of populist parties on the one hand, the rise of technocracy on the other (Caramani, 2017). The only thing these two forces hold in common is their rejection of mainstream party politics, their increasingly negative impact on such politics, and their deleterious effects on liberal democracy (Hobolt, 2015; Kriesi, 2016; Hooghe and Marks, 2009). The danger, as Yascha Mounk argues, is that liberal democracies may end up either with illiberal democracies run by populist demagogues or undemocratic liberalisms governed by technocratic elites (Mounk, 2018).

In sum, the depoliticizing effects of the supranationalization of decision-making, together with the weakening of representative party institutions, offer equally powerful explanations for how and why populism has emerged as a major challenge to mainstream parties and politics. But again, the question is why, given that this has been a long-term process, aggrieved citizens





**All forms of populism are expressions of discontent by those who feel dispossessed and are given voice by leaders whose discourses of dissent resonate with “the people’s” angry reactions against the status quo**

Torn and overlapping posters of the two presidential candidates in the French elections of 2017: the National Front's Marine Le Pen and En Marche's Emmanuel Macron







did not vote for populist parties on the right-wing extremes sooner. Cas Mudde suggests this may be a problem on the supply-side, that is, the absence of charismatic leaders attractive to the general voter for whom to vote (Mudde, 2017, p. 615) (despite the “coterie charisma” of some leaders felt by hard-core activists; Eatwell, 2017) But, if so, then the further question is why such populist leaders—some new but many still around after many years—have taken the world by storm only now.

## **Populist leaders articulate many more anti-system complaints about what is wrong than spell out proposals about how to fix it, at least until they gain access to power, at which point they may either row back or fast-forward on anti-liberal policies**

But in order to answer this question, we need to focus in on populism itself. Up to this point, we have looked at the sources of populist discontent by delving deeply into the causes of citizen’s discontent in three different areas—economic, social, and political. By focusing on the sources of the problem, the discussion tends to take populism as a given. Only by taking the ideas and discourse of populist movements and leaders seriously, however, can we come closer to understanding why populist forces have been able to exploit the current rise in citizen discontent for their own purposes.

### **Conceptualizing Populism and Its Effects**

Public and scholarly interest in the development of populism has spawned a veritable cottage industry of books and articles on the topic. Conceptually, scholars have provided important insights into the nature and scope of populism in Europe and America (for example, Mudde and Kalwasser, 2012; Müller, 2016; Judis, 2016; Mudde, 2017). Empirically, analysts have charted the rise of populism on the extremes of the left and the right, although the large majority are focused on the anti-immigrant, Eurosceptic, anti-euro, and anti-EU parties of the far right (for example, Kriesi, 2014, 2016; Mudde, 2017). Commentators have additionally shown that the problems generated by populism can be seen not just in the policy proposals that go counter to long-agreed principles of human rights, democratic processes, and the liberal world order but also in the new “uncivil” language of politics (Mutz, 2015; Thompson, 2016), the politics of “bullshit,” and the dangers of “fake news” circulating via the media to create a “post-truth” world (Frankfurt, 2005; Ball 2017; D’Ancona, 2017).

The high number and wide range of such works suggests that there is no one agreed-upon approach to understanding populism but many possible, most with negative connotations. Some take us all the way back to Richard Hofstadter’s depiction in the 1960s of “agitators with paranoid tendencies” (Hofstadter, 1964). Although that purely negative view of populism can be critiqued, in particular by differentiating left-wing from right-wing versions, all populism has one thing in common. It is the expression of discontent by those who feel dispossessed, given voice by leaders whose discourses of dissent resonate with “the people’s” angry reactions against the status quo. But beyond this, populism can follow many different avenues, depending upon the political, social, historical, institutional, and cultural context.



In taking account of this complexity, we can identify four key features of populism: first, populist leaders claim sole representation of “the people” against elites and other “threats.” Second, they engage in all-out assaults on expertise and unbiased “facts” and truth with “uncivil” language and “incivil” conduct that constitute a challenge to liberal tolerance and commitment to impartial information and scientific knowledge. Third, they get their messages out through new strategies of communication, facilitated by the new social media such as Twitter feeds and Facebook as well as the traditional broadcast and print media. And fourth, they articulate many more anti-system complaints about what is wrong than spell out proposals about how to fix it at least until they gain access to power, at which point they may either row back or fast-forward on anti-liberal policies.

**Populist Leaders’ Style of Discourse** Much attention in the literature on populism focuses on the first characteristic of populism, the appeals to “the people” by leaders whose discourses blame “corrupt” elites and unfair institutions for all their problems while enumerating a wide range of threats to national well-being, however that may be construed (Canovan, 1999; Weyland, 2001; Albertazzi and Mueller, 2017; Mudde, 2005, 2017). Most recent theoretical analyses of populism portray such discursive leadership as a danger for liberal democracy. Jan-Werner Müller, for example, defines populism rather narrowly as a dangerous anti-elitist, anti-democratic, and anti-pluralist political philosophy, in which leaders claim an exclusive representation of “the people”—with only some of the people counting as the “true people” for whom populist leaders claim to speak in the name of the people as a whole (Müller, 2016). This definition is close to that of Pierre André Taguieff, in his classic study of the National Front as a “national-populist party” in which the discourse of the demagogic leader is defined by a rhetoric that identifies with “the people,” claiming that their ideas are his, his ideas are theirs, with no concern for the truth, but rather to persuade through propagandistic formulas (Taguieff, 1984; see also discussion in Jäger, 2018).

**According to Laclau, populism is identifiable by its conceptual anchor, which stands as a universal representation for all other demands to which it is seen as equivalent. Examples might be slogans such as Brexit supporters’ “Take back control”**

A similar such approach from another philosophical tradition is that of Ernesto Laclau (2005, p. 39). He argues that populism is identifiable not so much by the contents or even the identification of an enemy as by its conceptual anchor (“empty signifier”), which stands as a universal representation for all other demands to which it is seen as equivalent (see also Panizza, 2005). Examples might be a general issue such as “globalization,” or phrases or slogans such as Brexit supporters’ “Take back control” and Donald Trump’s “Make America Great Again” (Schmidt, 2017). However, it could even consist of a string of words that indicate a particular set of values, as in the speech by Italian Interior Minister and head of the League, Matteo Salvini, at a rally in Pontida, who declared: “Next year’s [EP] election will be a referendum between the Europe of the elites, banks, finance, mass migration and precariousness versus the Europe of peoples, work, tranquility, family and future” (*Politico*, July 19, 2018).

For many, populism is an unqualified negative phenomenon: anti-democratic, anti-pluralist, and moralistic in extremely dangerous ways. This is particularly the case where the focus





Fifty-two year old Wink Watson of Lincoln, England, celebrates the result of the Brexit referendum outside the Britannia Pub on June 25, 2016





is on the rise of the “new populism” of extreme-right parties and their links to xenophobic nationalist ideas (for example, Taggart, 2017). These include far-right parties with reasonable longevity, such as France’s National Front (now National Rally), Austria’s Freedom Party, the Danish People’s Party, and the Dutch Party for Freedom (Elinas, 2010; Mudde, 2017); relative newcomers such as the Finns Party (formerly True Finns), the Sweden Democrats, and the Alternative for Germany (AfD) in northern Europe; as well as the variety of populist parties, new and old, across central and eastern Europe, including the Czech Republic and Slovakia (Minkenberg, 2002; Mudde, 2005); along with, of course, the illiberal governments of Hungary and Poland (Kelemen, 2017).

For others, populism can have a more positive side to it. This includes the left-wing populist governments of Latin America (especially in the 1990s and early 2000s) and the inclusionary populisms of southern Europe, most notably in Spain and Greece (Weyland, 2001; Panizza, 2005; Mudde and Kaltwasser, 2012). As some philosophers of the left such as Chantal Mouffe have argued (Mouffe, 2018), and many figures in the radical left political formations themselves (for example, Spain’s Podemos and France’s France Insoumise) have stressed, some radical left parties embrace the term populism as a technique for acquiring power. They see this as representing the only forceful and effective alternative on the left to the “surrender by consensus” carried out by a discredited social-democracy transformed by the Third Way.

## **Populism’s positive effects include giving voice to underrepresented groups, mobilizing and representing excluded sections of society, and increasing democratic accountability by raising issues ignored or pushed aside by the mainstream parties**

Populism’s positive effects include giving voice to underrepresented groups, mobilizing and representing excluded sections of society, and increasing democratic accountability by raising issues ignored or pushed aside by the mainstream parties. The extremes on the left in particular, by mobilizing on bases of social justice and human rights as well as against the inequalities caused by the increasing predominance of financial capitalism and its accompanying booms and busts, or by the lack of progressive taxation, can serve as a positive pull on mainstream parties—on the right as much as the left. The Occupy Movement is a case in point. However, there are many fewer extreme-left parties with a significant popular following than extreme-right parties, and they are often in EU countries that have less political pull or economic weight, in particular those which were subject to formal conditionality for bailouts during the euro crisis (that is, Greece; Vasilopoulou, 2018) or informal conditionality (most notably Spain). On balance, parties of the extreme right are the ones that appear to have exerted the most influence on political debates and the policy agenda so far, by pulling center-right mainstream parties closer to their positions, especially with regard to opposition to immigration and freedom of movement or minority rights.

The existence of different kinds of populist movements on a spectrum from left to right, whatever their relative strength, thus suggests that populism is more than just a discursive style with an anti-elite message. Although the style of populists may be similar—such as speaking in the name of the people against elites—the content does matter. If it is more progressive and inclusive, it can exert a positive influence on mainstream parties that serves



to reinforce liberal democracy. If more regressive and xenophobic, it can exert a negative influence. All populists are not the same, even if their styles may be similar. Ideological divides of the left and right remain of great importance, as a recent Pew study of citizens' support for populist versus mainstream parties of the left, right, and center concludes (Simmons et al., 2018).

**Populist Post-Truth** The next characteristic of populism in our list involves valuing personal experiences over knowledge and technical expertise. Populists tend to discredit experts, intellectuals, and those who have traditionally claimed to rely on “facts” and truth. This fight against experts is also at the origins of the many discussions of post-truth and fake news, both populists' accusations against mainstream news outlets of fake news any time the truth gets in their way and populists' own dissemination of fake news through social media as well as the traditional media (Ball, 2017; D'Ancona, 2017). Note, however, that this approach seems to apply much more to contemporary right populists than left populists.

Populists' contestation of expertise refers to the fact that they are prone to engage in the negation of the scientific/academic knowledge used by established political parties and generate their own “alternative” facts and sources of expertise, often by valuing personal experiences over “technocratic” expertise. To take but one example, Hungary's Jobbik has its own “institutes” that hybridize uncontested statistical facts on immigration with political myths from conspiracy theories lifted from anonymous producers on YouTube.

**Populists tend to discredit experts, intellectuals, and those who have traditionally claimed to rely on “facts” and truth. This fight is also at the origins of the many discussions of post-truth and fake news, both populists' accusations against mainstream news outlets of fake news any time the truth gets in their way and populists' own dissemination of fake news through social media as well as the traditional media**

The problem with this blurring of the lines between fact and fiction, as psychologists have pointed out, is that it undermines people's very sense of truth or falsehood, as lies repeated many times are believed as “true” even when people know they are not. Here, we can learn a lot from the work of psychologists who focus on the ways in which framing and *heuristics* can affect people's perceptions (for example, Kahneman, 2011; Lackoff, 2014), including when exaggeration or hyperbole, say, of the number of migrants entering the EU or the cost of the EU per day to the UK, leaves the impression on listeners that a very large number is involved, even if not as high as alleged. Even speech patterns, such as incomplete sentences and repetitions, can serve as effective discursive mechanisms to reinforce a message, whether by creating a sense of intimacy as audiences complete the sentence in their heads, or appealing to unconscious cognitive mechanisms that serve to reinforce people's acceptance of what is said, even (or especially) when they are lies and exaggerations (Lackoff, 2016—see discussion in Schmidt, 2017).

This kind of post-truth approach to the world is part and parcel of the combative “uncivil” language and style of discursive interaction, in which bullying, shouting, and blatantly

violating the rules of “political correctness” through intolerant language contribute to the sense that it is not just what you say but how assertively you say it, regardless of the validity of the claims, that counts. The danger here is that it undermines the very values—of tolerance, fairness, and even-handed reporting—that have been at the basis of liberal democracy since the postwar period. As Diane Mutz argues, the incivility in the mass media, in particular on confrontation “in-your-face” news programs, is particularly detrimental to facilitating respect for oppositional political viewpoints and to citizens’ levels of trust in politicians and the political process (Mutz, 2015).



**Political Coordination through New Social Media** Contemporary populism also goes hand in hand with the new ways in which populists have learned to use new social media to circulate their messages and broaden their networks of support and resource base. Indeed, new media have been invaluable to populists’ creation of networks of dissent. Facebook posts alone, for example, create echo chambers of support, in particular because large numbers of people get their news (fake as well as real) from their “friends” sharing posts. Populists rely more on new media (for example, YouTube and blogs) and social media (for example, Twitter and Facebook) than traditional parties do. For example, in Spain, Podemos faced down the hostility of newspapers and television outlets with extreme reliance on hyperactive Facebook posts and YouTube channel streaming. Social media facilitates the discovery of like-minded people across the country and the world—enabling populist activists and parties to exponentially increase the number of their “followers” and potential supporters. Transnational networks of communication enable the spread of populist ideas, reinforcing anger and anti-establishment sentiment. Crucially, however, this happens not only virtually but also “in the flesh,” for example, when extreme-right leaders meet in Europe to set strategies for EP elections or parliamentary groupings. A recent case in point is when President Trump’s “organic intellectual,” Steve Bannon, traveled throughout Europe to meet with and support other populist leaders in their electoral battles, such as Nigel Farage and Marine Le Pen, and plans to set up a foundation to provide advice and financial support.

Populism finds support from activists and social movements on both the left and the right. While it is commonly assumed that the activist networks are primarily engaged in left-leaning causes, right-wing networks have also been active. In the US, the Tea Party is the clearest example, managing to unseat enough incumbents in primaries and to win elections so as to transform the Republican party (Skocpol and Williamson, 2012). In the UK, the United Kingdom Independence Party (UKIP) was able to set the agenda for the Conservative party and, ultimately, the entire nation through the referendum on UK exit from the EU. In some European cases, as in Denmark (Rydgren, 2004) and Germany (Berbuir et al., 2015), social movements have been instrumental in propelling and normalizing right-wing populism (see also Bale et al., 2010). All of this said, populism has also been useful to left-wing activists seeking to enliven their support base (March and Mudde, 2005). Bernie Sanders in the Democratic primaries of 2016 has sometimes been called a populist because of his ability to energize youth via social media, despite or perhaps because of promises that mainstream Democrats claimed were unrealistic.

**Political Communication via the Traditional Media** The dissemination of populist views does not come just from new social media that create new channels of coordination by activists’ networks, however. Populists have also exploited the old media to get their messages out beyond their “true believers” to the more general public. While Twitter feeds provide a post-modern way for populist leaders to speak directly to “the people,” the traditional media also help





spread their messages of distrust of mainstream parties and politics as well as the media itself. As linguist Ruth Wodak shows, with the “politics of fear,” right-wing populist parties have gone from fringe voices to persuasive political actors who set the agenda and frame media debates via the normalization of nationalistic, xenophobic, racist, and anti-semitic rhetoric (Wodak, 2015). That said, dissemination of the populist discourse does have its limits, since some things do not translate—as when US alt-right activists sought to use “Freddie the Frog” to reinforce extreme-right sentiment in France in the run-up to the presidential election—not realizing that “frog” has long been a negative stereotype applied to the French, and therefore would not resonate.

In many countries, the traditional media has become so fragmented that people listen to different news programs with completely different slants on the news. And, here again, it is mostly the extreme right that largely wins over the left with regard to broadcasting presence, whether in terms of talk radio or cable news, whether Radio Maria in Poland or Fox News in the US. Moreover, even the mainstream press and TV conspires to favor the extremes on the right, if only inadvertently. They magnify the audience of populist leaders whose political “incorrect” tweets become the news story of the day, or they reinforce right-of-center messages when in efforts to appear “balanced” they bring on someone from the extreme right and someone from the center—without any airtime for the extreme left (Baldwin, 2018). Naturally, where the populists are in government and control the traditional media, then the populist message is the main one heard—as is the case of Hungary, but arguably even in Italy under Berlusconi’s more benign version of populism.

Media communication has also changed in ways that benefit populist messaging. The short news cycles, combined with the push to speak in thirty-second sound-bites, privileges simpler messages that “sell,” and this in turn favors populists with their simple “solutions” to complex problems, easy to articulate without explanation: such as “build a wall” to solve the immigration problem, reverse free trade to protect jobs in the country, and so forth. It takes much longer for mainstream leaders to explain why certain kinds of policies are in place, and often these explanations are complex and boring, especially when compared to the snappy slogans of the populists.

## **Populist discourse focuses more on listing grievances and injustices than on laying out policy prescriptions and detailed policy programs. As such, it tends to work best for populists in opposition**

This “mediatization” of political communication generally poses significant problems for mainstream party politics and government, primarily by undermining mainstream party control of the public sphere and mainstream parties’ ability to set the political agenda. Beyond the fact that many other non- or anti-establishment voices are now heard through a multiplicity of channels, mainstream leaders have created their own problems as a result of their own more populist styles of communication, while the media have only added to these through their tendency to focus on leaders’ personality traits while turning the news into entertainment. Beyond this, the social media, social movements, and out-groups have also been increasingly subverting the political agenda-setting function of political parties (Caramani, 2017). Political communication, then, in the dissemination of populist ideas and



Occupy Wall Street supporters attending a street concert by the group Rage Against the Machine at Foley Square, New York, during the celebration of the first anniversary of the citizens' movement, which was born on September 17, 2011



discourse through the “bullshit” of fake news and post-truth in a fragmented media landscape is another key element of populism today.



**Connecting Populist Discourses to Actions** Our last dimension of populism is leaders’ tendency to focus more on denouncing the status quo than suggesting remedies, until they gain political power. Populism most often entails, as mentioned above, an ideologically thin discourse characterized more by the ardent expression of resentment than by the consistency of the programs (also termed a “thin-centered ideology”—Mudde and Kaltwasser, 2012, p. 8). The populist discourse is therefore more likely to focus on listing grievances and injustices rather than laying out policy prescriptions and detailed policy programs. As such, this tends to work best for populists in opposition. Being in government has long entailed compromise or even turn-around on cherished policies (Mudde, 2017)—as in the case of the left-wing Syriza in Greece. But recently, such turn-arounds have become less frequent.

As more and more populist parties have been joining mainstream party coalitions (for example, Austria), or even governing on their own (in Italy, Hungary, and Poland), they have been designing and implementing policy agendas that put into effect their anti-liberal ideas, often with only the courts to safeguard the rule of law. Moreover, as the chances of election are increasing for populists across Europe, all such parties have become more specific about their policies and programs. And they do this even when (or especially when) such policies cannot easily be implemented under the existing political order because they violate assumptions about sound economics (for example, promising a high guaranteed income *and* a flat tax—as in the program of the new populist coalition government in Italy) or liberal politics (for example, expelling refugees—the pledge of all right-wing populist parties).

So, exactly what are the potential dangers when populists gain power? David Art has argued that the political strategy of “tamed power,” by bringing populists into government to force them to take on their responsibilities via compromise, can backfire, by “normalizing” their ideas and thereby opening the way for illiberal ideas to gain sway in liberal democracies (Art, 2006; see also Mudde, 2017). Müller goes farther, to contend that rather than encouraging a more participative democracy, populists in power will “engage in occupying the state, mass clientelism and corruption, and the suppression of anything like a critical civil society” (Müller, 2016, p. 102). Steve Levitsky and Daniel Ziblatt echo this analysis, insisting that “democracies die” at the hands of elected populist leaders who then subvert the very democratic processes that brought them to power (Levitsky and Ziblatt, 2018). But even short of populist victory, when populists are not in power *yet*, the dangers also come from contagion. Mainstream leaders are themselves increasingly guilty of introducing populist styles of discourse into normal politics, with the “electoralism” of political parties’ increasing emphasis on short-term electoral goals while responding to the public mood as gauged through polling instruments (Caramani, 2017). This suggests that it is not enough to track leaders’ discourses and the ways in which their ideas circulate. We also need to see whether and/or how they influence liberal democracies.

## Conclusion

We are left with a number of questions. Is this a moment of great transformation, in which a new paradigm will emerge out of the ashes of the liberal order, with neoliberal economics, social liberalism, and political liberalism succumbing to the closing of borders to immigrants, rising protectionism, social conservatism, and illiberal democracy (itself an oxymoron)? Will

the more balanced and tolerant institutional commitments of political liberalism prevail, along with a perhaps modified economic liberalism in which open borders and free trade are moderated by more attention to those left behind? For the moment, we cannot know. What we do know is that when populist leaders gain power, they try to make good on their promises, to the detriment of the liberal democratic consensus.

So, what is the alternative? The big question for progressives who seek to maintain liberal democracies is how to counter the populist upsurge with innovative ideas that go beyond neoliberal economics while promoting a renewal of democracy and a more egalitarian society. But this requires not just workable ideas that can provide real solutions to the wide range of problems related to economics, politics, and society. It also demands political leaders with persuasive discourses that can resonate with an increasingly discontented electorate, more and more open to the sirens of populism. For the moment, we continue to wait not so much for the ideas—in many ways we know what they are—but for the discourse of new political leaders able to convey progressive ideas in uplifting ways that offer new visions of the future able to heal the schisms on which the populists have long thrived. Without this, hopes of any “New Enlightenment” will be dashed on the shoals of illiberalism.





## Select Bibliography

- Albertazzi, Daniele, and Mueller, Sean. 2017. "Populism and liberal democracy." In *The Populist Radical Right: A Reader*, Cas Mudde (ed.). London: Routledge.
- Art, David. 2006. *The Politics of the Nazi Past in Germany and Austria*. New York: Cambridge University Press.
- Baldwin, Tom. 2018. *Ctrl Alt Delete: How Politics and the Media Crashed our Democracy*. London: Hurst and Compan.
- Bale, Tim, Green-Pedersen, Christoffer, Krouwel, Andrea, Luther, Kurt Richard, and Sitter, Nick. 2010. "If you can't beat them, join them? Explaining social democratic responses to the challenge from the populist radical right in Western Europe." *Political Studies* 58(3): 410–426.
- Ball, James. 2017. *Post-Truth: How Bullshit Conquered the World*. London: Biteback Publishing.
- Berbair, Nicole, Lewandowsky, Marcel, and Siri, Jasmin. 2015. "The AfD and its sympathisers: Finally a right-wing populist movement in Germany?" *German Politics* 24(2): 154–178.
- Berezin, Mabel. 2009. *Illiberal Politics in Neoliberal Times*. New York: Cambridge University Press.
- Berman, Sheri. 2018. "Populism and the future of liberal democracy in the West." Presentation at the Center for the Study of Europe, Boston University (Boston, Sept. 20, 2018).
- Canovan, Margaret. 1999. "Trust the people! Populism and the two faces of democracy." *Political Studies* 47(1): 2–16.
- Caramani, Daniele. 2017. "Will vs. reason: Populist and technocratic challenges to representative democracy." *American Political Science Review* 111(1): 54–67.
- D'Ancona, Matthew. 2017. *Post Truth: The New War on Truth and How to Fight Back*. London: Ebury Press.
- De Wilde, Piete, and Zürn, Michael. 2012. "Can the politicisation of European integration be reversed?" *Journal of Common Market Studies* 50(1): 137–153.
- Eatwell, Roger. 2017. "The rebirth of right-wing charisma?" In *The Populist Radical Right: A Reader*, Cas Mudde (ed.). London: Routledge.
- Eberstadt, Nicholas. 2016. *Men Without Work*. West Conshohocken, PA: Templeton Press.
- Eichengreen, Barry. 2018. *The Populist Temptation: Economic Grievance and Political Reaction in the Modern Era*. New York: Oxford University Press.
- Elinas, Antonis. 2010. *The Media and The Far Right in Western Europe: Playing the Nationalist Card*. New York: Cambridge University Press.
- Fawcett, P., and Marsh, D. 2014. "Depoliticisation, governance and political participation." *Policy & Politics* (special issue) 42(2): 171–188.
- Frankfurt, Harry. 2005. *On Bullshit*. Princeton: Princeton University Press.
- Gidron, Noam, and Hall, Peter A. 2017. "The politics of social status: economic and cultural roots of the populist right." *The British Journal of Sociology* 68(1): 57–68.
- Hacker, Jacob. 2006. *The Great Risk Shift: The Assault on American Jobs, Families, Health Care, and Retirement, and How You Can Fight Back*. New York: Oxford.
- Hacker, Jacob S., and Pierson, Paul. 2010. *Winner-Take-All Politics: How Washington Made the Rich Richer—And Turned Its Back on the Middle Class*. New York: Simon and Schuster.
- Hay, Colin, and Wincott, Daniel. 2012. *The Political Economy of European Welfare Capitalism*. Basingstoke: Palgrave Macmillan.
- Hemerijck, Anton. 2013. *Changing Welfare States*. Oxford: Oxford University Press.
- Hobolt, Sara. 2015. "Public attitudes toward the Eurozone crisis." In *Democratic Politics in a European Union under Stress*, O. Cramme and S. Hobolt (eds.). Oxford: Oxford University Press.
- Hochschild, Arlie Russell. 2016. *Strangers in Their Own Land*. New York: New Press.
- Hochschild, Jennifer, and Mollenkott, John H. 2009. *Bringing Outsiders In: Transatlantic Perspectives on Immigrant Political Incorporation*. Ithaca: Cornell.
- Hofstadter, Richard. 1964. *The Paranoid Style in American Politics and Other Essays*. New York: Alfred A. Knopf.
- Hooghe, L., and Marks, G. 2009. "A postfunctionalist theory of European integration: From permissive consensus to constraining dissensus." *British Journal of Political Science* 39 :1: 1–23.
- Inglehart, Ronald, and Norris, Pippa. 2016. "Trump, Brexit and the rise of populism: Economic have-nots and cultural backlash." Paper prepared for the Annual meeting of the American Political Science Association (Philadelphia, September 1–4).
- Jäger, Anton. 2018. "The myth of populism." *Jacobin* (January 3) <https://www.jacobinmag.com/2018/01/populism-douglas-hofstadter-donald-trump-democracy>.
- Judis, John B. 2016. *The Populist Explosion: How the Great Recession Transformed American and European Politics*. New York: Columbia Global Reports.
- Kahneman, Daniel. 2011. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kelemen, Daniel. 2017. "Europe's authoritarian equilibrium." *Foreign Affairs* <https://www.foreignaffairs.com/print/1121678>.
- Kriesi, Hans-Peter. 2014. "The populist challenge." *West European Politics* 37(2): 379–399.
- Kriesi, H. 2016. "The politicization of European integration." *Journal of Common Market Studies* 54 (annual review): 32–47.
- Kriesi, H., Grande, E., Dolezal, M., Helbling, M., Höglinger, D., Hutter, S., and Wueest, B. 2012. *Political Conflict in Western Europe*. Cambridge: Cambridge University Press.
- Lackoff, George. 2014. *Don't Think of an Elephant? Know Your Values and Frame the Debate*. White River Junction, VT: Chelsea Green Publishing.
- Lackoff, George. 2016. "Understanding Trump's use of language." *Social Europe*, August 23.
- Laclau, Ernesto. 2005. *On Populist Reason*. London: Verso.
- Levitsky, Steven, and Ziblatt, Daniel. 2018. *How Democracies Die*. New York: Viking.
- Lilla, Mark. 2017. *The Once and Future Liberal*. New York: Harper.
- Mair, Peter. 2013. *Ruling the Void: The Hollowing of Western Democracy*. London: Verso.
- March, Luke, and Mudde, Cas. 2005. "What's left of the radical left? The European radical left after 1989: Decline and mutation." *Comparative European Politics* 3(1): 23–49.
- McClaren, Lauren. 2012. "The cultural divide in Europe: Migration, multiculturalism, political trust." *World Politics* 64(2): 199–241.
- Minkenberg, Michael. 2002. "The radical right in postsocialist central and eastern Europe: Comparative observations and interpretations." *East European Politics and Societies* 16(2): 335–362.
- Mouffe, Chantal. 2018. *For a Left Populism*. London: Verso Books.
- Mounk, Yascha. 2018. *The People vs. Democracy: Why Our Freedom Is in Danger & How to Save It*. Cambridge: Harvard University Press.
- Mudde, Cas (ed.). 2005. *Racist Extremism in Central and Eastern Europe*. London: Routledge.
- Mudde, Cas. 2017. "Conclusion: Studying populist radical right parties and politics in the twenty-first century." In *The Populist Radical Right: A Reader*, Cas Mudde (ed.). London: Routledge.
- Mudde, Cas, and Kaltwasser, Cristobal Rovira. 2012. *Populism in Europe and the Americas: Threat or Corrective to Democracy*. Cambridge: Cambridge University.
- Müller, Jan-Werner. 2016. *What Is Populism?* Philadelphia: University of Pennsylvania Press.
- Mutz, Diane. 2015. *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton: Princeton University Press.
- Panizza, Francisco. 2005. "Introduction: Populism and the mirror of democracy." In *Populism and the Mirror of Democracy*. London: Verso.
- Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Belknap Press of Harvard University.
- Prosser, Thomas. 2016. "Insiders and outsiders on a European scale." *European Journal of Industrial Relations* online doi: 10.1177/0959680116668026.

- Rodrik, Dani. 2011. *The Globalization Paradox*. New York: Norton.
- Rosenbluth, Frances McCall, and Shapiro, Ian. 2018. *Responsible Parties: Saving Democracy from Itself*. New Haven: Yale University Press.
- Rydgren, Jens. 2004. "Explaining the emergence of radical right-wing populist parties: The case of Denmark." *West European Politics* 27(3): 474–502.
- Scharpf, Fritz W. 2014. "Political legitimacy in a non-optimal currency area." In *Democratic Politics in a European Union under Stress*, Olaf Cramme, and Sara B. Hobolt (eds.). Oxford: Oxford University Press.
- Schmidt, Vivien A. 2002. *The Futures of European Capitalism*. Oxford: Oxford University Press.
- Schmidt, Vivien A. 2006. *Democracy in Europe: The EU and National Politics*. Oxford: Oxford University Press.
- Schmidt, Vivien A. 2017. "Britain-out and Trump-in: A discursive institutionalist analysis of the British referendum on the EU and the US presidential election." *Review of International Political Economy* 24(2): 248–269.
- Schmidt, Vivien A., and Thatcher, Mark. 2013. "Introduction: The resilience of neo-liberal ideas." In *Resilient Liberalism in Europe's Political Economy*, V. Schmidt and M. Thatcher (eds.). Cambridge: Cambridge University Press.
- Simmons, Katie, Silver, Laura, Johnson, Courtney, Taylor, Kyle, and Wike, Richard. 2018. "In Western Europe, populist parties tap anti-establishment frustration but have little appeal across ideological divide." Pew Research Center (July 12) [http://assets.pewresearch.org/wp-content/uploads/sites/2/2018/07/12092128/Pew-Research-Center-Western-Europe-Political-Ideology-Report\\_2018-07-12.pdf](http://assets.pewresearch.org/wp-content/uploads/sites/2/2018/07/12092128/Pew-Research-Center-Western-Europe-Political-Ideology-Report_2018-07-12.pdf).
- Skocpol, Theda, and Williamson, Vanessa. 2012. *The Tea Party and the Remaking of American Conservatism*. New York: Oxford University Press.
- Standing, Guy. 2011. *The Precariat: The New Dangerous Class*. London: Bloomsbury.
- Stiglitz, Joseph. 2002. *Globalization and Its Discontents*. New York: Norton.
- Taggart, Paul. 2000. "Populism." Buckingham, UK, Open University Press.
- Taggart, Paul. 2017. "New populist parties in Western Europe." In *The Populist Radical Right: A Reader*, Cas Mudde (ed.). London: Routledge.
- Taguieff, Pierre André. 1984. "La rhétorique du national-populisme." *Mots* 9: 113–138.
- Thompson, Mark. 2016. *Enough Said: What's Gone Wrong with the Language of Politics?* New York: St. Martin's Press.
- Vasilopoulou, Sofia. 2018. "The party politics of Euroscepticism in times of crisis: The case of Greece." *Politics*. Online first at: <http://eprints.whiterose.ac.uk/128036/>.
- Weyland, Kurt. 2001. "Clarifying a contested concept: Populism in the study of Latin American politics." *Comparative Politics* 39(1): 1–34.
- Wodak, Ruth. 2015. *The Politics of Fear: What Right Wing Populist Discourses Mean*. London: Sage.



**Diana Owen**  
Georgetown University

Dr. Diana Owen is Professor in the Communication, Culture, and Technology graduate program and former Director of American Studies at Georgetown University. A political scientist, her research explores new political media and citizenship education in the digital age. She is the author of *Media Messages in American Presidential Elections*, *New Media and American Politics* (with Richard Davis), and *American Government and Politics in the Information Age* (with David Paletz and Timothy Cook). She is co-editor of *The Internet and Politics: Citizens, Voters, and Activists*; *Making a Difference: The Internet and Elections in Comparative Perspective*; and *Internet Election Campaigns in the United States, Japan, Korea, and Taiwan*. She earned grants from the Pew Charitable Trusts, the Center for Civic Education, Storyful/News Corp, Google, and the US Department of Education. Dr. Owen was an American Political Science Association Congressional Media Fellow and received the Daniel Roselle Award from the Middle States Council for the Social Studies.

Recommended book: *American Government and Politics in the Information Age*, David Paletz, Diana Owen, and Timothy Cook, FlatWorld Knowledge Press, 2018.

**The American media system has undergone significant transformations since the advent of new media in the late 1980s. During the past decade, social media have become powerful political tools in campaigns and governing. This article explores three major trends related to the rise of social media that are relevant for democratic politics in the United States. First, there is a major shift in how and where people get political information, as more people turn to digital sources and abandon television news. Next, the emergence of the political “Twitterverse,” which has become a locus of communication between politicians, citizens, and the press, has coarsened political discourse, fostered “rule by tweet,” and advanced the spread of misinformation. Finally, the disappearance of local news outlets and the resulting increase in “news deserts” has allowed social-media messages to become a primary source of information in places where Donald Trump’s support is most robust.**



The political media system in the United States has undergone massive transformations over the past three decades. The scope of these new media developments is vast, encompassing both legacy sources as well as entirely novel communication platforms made possible by emerging technologies. The new media era began with the infotainment trend in the 1980s when television talk shows, talk radio, and tabloid papers took on enhanced political roles. Changes became more radical when the Internet emerged as a delivery system for political content in the 1990s. Digital technology first supported platforms where users could access static documents and brochures, but soon hosted sites with interactive features. The public gained greater political agency through technological affordances that allowed them to react to political events and issues, communicate directly to candidates and political leaders, contribute original news, images, videos, and political content, and engage in political activities, such as working on behalf of candidates, raising funds, and organizing protests. At the same time, journalists acquired pioneering mechanisms for reporting stories and reaching audiences. Politicians amassed new ways of conveying messages to the public, other elites, and the press, influencing constituents' opinions, recruiting volunteers and donors, and mobilizing voters (Davis and Owen, 1998; Owen, 2017a).

The evolution of social media, like Facebook, Twitter, and YouTube, from platforms facilitating networks among friends to powerful political tools has been an especially momentous development. The political role of social media in American politics was established during the 2008 presidential election. Democratic presidential candidate Barack Obama's social-media strategy revolutionized campaigning by altering the structure of political organizing. Obama's campaign took on the characteristics of a social movement with strong digital grassroots mobilization (Bimber, 2014). The campaign exploited the networking, collaborating, and community-building potential of social media. It used social media to make personalized appeals to voters aided by data analytics that guided targeted messaging. Voters created and amplified messages about the candidates without going through formal campaign organizations or political parties (Stromer-Galley, 2016). The most popular viral videos in the 2008 campaign, BarelyPolitical.com's "Obama Girl" and will.i.am's "Yes, We Can," were produced independently and attracted millions of viewers (Wallsten, 2010). In this unique election, the calculated strategies of Obama's official campaign organization were aided by the spontaneous innovation of voters themselves. Going forward, campaigns—including Obama's 2012 committee—would work hard to curtail outside efforts and exert more control over the campaign-media process (Stromer-Galley, 2016).

## **The new media era began with the infotainment trend in the 1980s when television talk shows, talk radio, and tabloid papers took on enhanced political roles**

Since then, social media's political function in campaigns, government, and political movements, as well as their role in the news media ecosystem, has rapidly broadened in reach, consequence, and complexity. As political scientist Bruce Bimber points out: "The exercise of power and the configuration of advantage and dominance in democracy are linked to technological change" (2014: 130). Who controls, consumes, and distributes information is largely determined by who is best able to navigate digital technology. Social media have emerged as essential intermediaries that political and media actors use to assert influence. Political leaders have appropriated social media effectively to achieve political ends, ever-more fre-





quently pushing the boundaries of discursive action to extremes. Donald Trump's brash, often reckless, use of Twitter has enabled him to communicate directly to the public, stage-manage his political allies and detractors, and control the news agenda. Aided by social media, he has exceeded the ability of his modern-day presidential predecessors to achieve these ends. Social-media platforms facilitate the creation and sustenance of ad hoc groups, including those on the alt-right and far left of the political spectrum. These factors have encouraged the ubiquitous spread of false information that threatens to undermine democratic governance that relies on citizens' access to quality information for decision-making.

Social media have low barriers to entry and offer expanded opportunities for mass political engagement. They have centralized access to information and have made it easier for the online population to monitor politics. Growing numbers of people are using social media to engage in discussions and share messages within their social networks (Owen, 2017b). Effective use of social media has contributed to the success of social movements and political protests by promoting unifying messages and facilitating logistics (Jost et al., 2018). The #MeToo movement became a global phenomenon as a result of social media spreading the word. Actress Alyssa Milano sent out a tweet encouraging women who had been sexually harassed or assaulted to use the #MeToo hashtag in their social-media feed. Within twenty-four hours, 4.7 million people on Facebook and nearly one million on Twitter had used the hashtag. The number grew to over eighty-five million users in eighty-five countries on Facebook in forty-five days (Sayej, 2017).

Still, there are indications that elite political actors have increasingly attempted to shape, even restrict, the public's digital influence in the political sphere. Since 2008, parties and campaign organizations have sought to hyper-manage voters' digital engagement in elections by channeling their involvement through official Web sites and social-media platforms. They have controlled voters' access to information by microtargeting messages based on users' personal data, political proclivities, and consumer preferences derived from their social-media accounts. Further, a small number of companies, notably Google and Facebook, have inordinate power over the way that people spend their time and money online. Their ability to attract and maintain audiences undercuts the ability of small firms, local news outlets, and individuals to stake out their place in the digital market (Hindman, 2018).

## **The political role of social media in American politics was established during the 2008 presidential election. Democratic presidential candidate Barack Obama's social-media strategy revolutionized campaigning by taking on the characteristics of a social movement with strong digital grass-roots mobilization**

This article will focus on three major trends related to the rise of social media over the past decade that have particular significance for democratic politics and governance in the United States. First, the shift in audience preferences away from traditional mass media to digital sources has changed how people follow politics and the type of information they access. Many people are now getting news from their social-media feeds which contributes to rampant political insularity, polarization, and incivility. Next, the emergence of the political "Twitterverse" has fundamentally altered the way that politicians, citizens, and the press convey information, including messages of significant import to the nation. The "Twitterverse" is



comprised of the users of the microblogging platform as well as those exposed to its content when it is disseminated through other media, such as twenty-four-hour news channels. Twitter and other social media in the age of Trump have advanced the proliferation of disinformation, misinformation, “alternative facts,” and “fake news.” Importantly, Donald Trump’s presidency has ushered in an era of “rule by tweet,” as politicians make key pronouncements and conduct government business through Twitter. Finally, the spread of “news deserts”—places where local news outlets have disappeared—has compromised the institutional media’s ability to check false facts disseminated by social media, hyper-partisan sources, and bots propagating computational propaganda.

### Shifting Audience Media Preferences

The number of options available to news consumers has grown dramatically as content from ever-increasing sources is distributed via print, television, radio, computers, tablets, and mobile devices. More Americans are seeking news and political information since the 2016 presidential election and the turbulent times that have followed than at other periods in the past decade. At the same time, seven in ten people have experienced news fatigue and feel worn out from the constant barrage of contentious stories that are reported daily (Gottfried and Barthel, 2018). This is not surprising when the breaking news within a single day in September 2018 featured Dr. Christine Blasey Ford’s testimony that she had been sexually assaulted by Judge Brett Kavanaugh, a nominee for the US Supreme Court; the possibility that Deputy Attorney General Rod Rosenstein, head of the investigation into Russian intervention in the 2016 presidential election, could be fired by President Donald Trump for suggesting in a private meeting that steps be taken to remove Trump from office; and comedian Bill Cosby being sentenced to prison for sexual misconduct and led away in handcuffs.

**Where Americans Get Their News** One of the most notable developments in the past decade has been the shift in where and how Americans get their news and information about politics. There has been a marked transition in audience preferences away from traditional media, especially television and print newspapers, to online news sources and, more recently, news apps for smartphones. Social media have become major sources of news for millions of Americans, who either get political information deliberately through subscriptions or accidentally come upon it in their newsfeed. Trends in the public’s media use become most apparent during periods of heightened political awareness, such as during political campaigns. Thus, Table 1 presents the percentage of adults who frequently used specific types of media to get information about the 2008, 2012, and 2016 presidential elections.

The excitement surrounding the 2008 election, largely attributed to the landmark candidacy of Barack Obama, coupled with digital breakthroughs in campaigning, caused regular online news use to escalate to 37% of the public from 9% in 2006. Attention to news online dwindled somewhat in 2012 to 31%, as voters were less interested in the 2012 presidential campaign where President Obama sought reelection against Republican challenger Mitt Romney. However, the public’s use of news delivered online and through apps surged to 43% during the 2016 election that pitted Democrat Hillary Clinton against Republican Donald Trump.

When cable, network, and local channels are considered together, television was the main source of campaign information throughout this period for a majority of the public. Network TV news had seen a precipitous decline in viewership prior to 2008, and its regular audience



University of Southern California (USC) students watching the real-time broadcast of Brett Kavanaugh's candidature for the US Supreme Court. Kavanaugh was accused of sexual abuse by Christine Blasey Ford (in the image at the left). September 27, 2018





remained consistently around 30% of the population across the three election cycles, then falling in 2017 to 26%. Cable TV news' popularity declined somewhat from around 40% in 2008 and 2012 to 31% in 2016 and 28% in 2017. Like network news, local TV news has dropped in popularity over the past two decades. This decline may in part be attributed to the disappearance of independent local news programs that have been replaced by shows operated by Sinclair Broadcast Group, a conservative media organization. At its peak, local news was viewed regularly by more than 70% of the population. Local news attracted the largest regular TV audience in 2008 (52%), dropped precipitously in 2012 (38%), climbed again in popularity in 2016 (46%), and fell to 37% in 2017 (Matsa, 2018).

## **In 2016, 57% of the public often got news on television compared to 38% who used online sources. From 2016 to 2017, television's regular audience had declined to 50% of the population, and the online news audience had grown to 43%**

In a relatively short period of time, the public's preference for online news has made significant gains on television news as a main source. In 2016, 57% of the public often got news on television compared to 38% who used online sources. From 2016 to 2017, television's regular audience had declined to 50% of the population, and the online news audience had grown to 43%. The nineteen-percentage point gap in favor of television news had closed to seven percentage points in a single year (Gottfried and Shearer, 2017). If current trends continue, online news may eclipse television news as the public's main source in the not-so-distant future.

As the public's preference for online news has surged, there has been a major decline in print newspaper readership. In 2008, 34% of the public regularly read a print paper. By 2012 the number had decreased to 23% and continued to fall to 20% in 2016 and 18% in 2017. Radio news has maintained a relatively stable audience share with occasional peaks when political news is especially compelling, such as during the 2008 presidential campaign when 35% of the public regularly tuned in. The radio news audience has remained at around 25% of the public since 2012. Talk radio, bolstered by the popularity of conservative hosts, such as Rush Limbaugh, and community radio reporting on local affairs consistently attracts around 10% to 12% of the population (Guo, 2015).

**Social Media as a News Source** The American public's use of social media increased rapidly in the period following the 2008 presidential election. Reliance on social media for news and political information has increased steadily over the past decade. According to the Pew Research Center, 68% of American adults in 2018 got news from social media at least occasionally, and 20% relied often on social media for news (Shearer and Matsa, 2018). Table 2 presents data from the Pew Research Center indicating the percentage of Americans who regularly used at least one social-media site like Facebook, Twitter, or LinkedIn over time. Few people were active on social media between 2005 and 2008. Even during the watershed 2008 campaign, only 21% of the public was on social media. By 2009, however, the number of people online had spiked to 42% as social media took hold in the political sphere in the run-up to the 2010 midterm elections. The Tea Party, a loosely-organized populist movement that ran candidates who successfully attained office, relied heavily on social media as an alternative to the



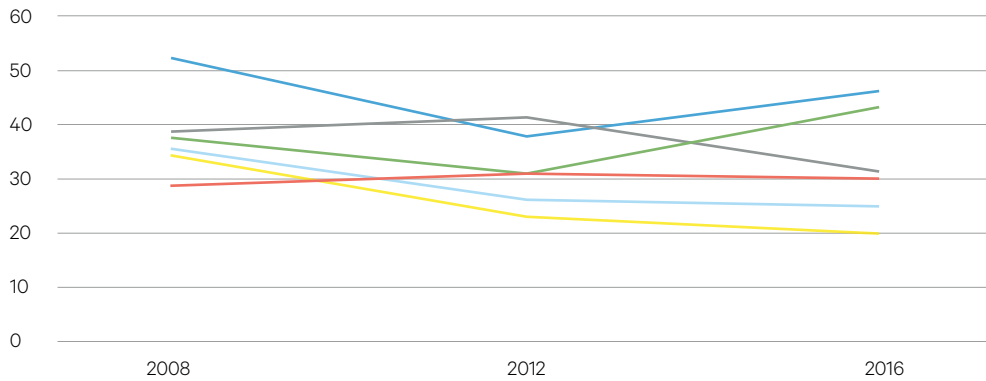


Table 1. Percentage of American adults using media often in the 2008, 2012, and 2016 presidential elections. (Source: Pew Research Center, data compiled by the author.)

— Cable television news  
— Network television news  
— Local television news  
— Print media  
— Radio  
— Internet and other online news

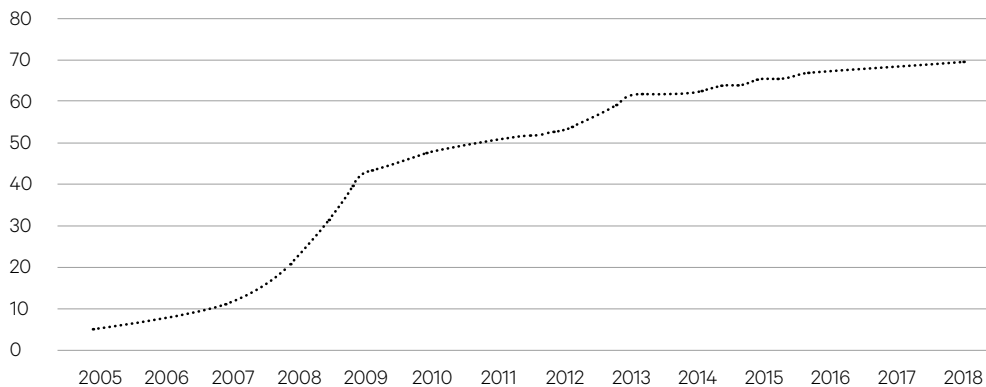


Table 2. Americans' social-media use 2005–18. (Source: Pew Research Center, Social-Media Fact Sheet, 2018.)

mainstream press which they regularly assailed. The mainstream press was compelled to cover the Tea Party's social-media pronouncements, which helped digital platforms to gain in popularity among voters. Social-media users who remained loyal to the Tea Party were prominent among supporters who actively worked on behalf of Donald Trump's campaign by mobilizing voters in their networks (Rohlinger and Bunnage, 2017). By 2013, over 60% of the public was using social media. The percentage of social-media users has leveled off at near 70% since the 2016 presidential election.

The shift in the public's media allegiances toward digital sources has rendered social media a far more viable and effective political tool. A decade ago, only the most interested and tech-savvy citizens used social media for politics. Young people were advantaged in their ability to leverage social media due to their facility with the technology and their fascination with the novelty of this approach to politics. The age gap in political social-media use has been



closing, as have differences based on other demographic characteristics, such as gender, race, education, and income (Smith and Anderson, 2018), which has altered politicians' approach to these platforms. In the past, elites employed social media primarily to set the agenda for mainstream media so that their messages could gain widespread attention. Today, political leaders not only engage social media to control the press agenda, they can also use these platforms effectively to cultivate their political base. In addition, elites use social media to communicate with one another in a public forum. In 2017, Donald Trump and North Korean president Kim Jong Un traded Twitter barbs that escalated tensions between the two nations over nuclear weapons. They exchanged personal insults, with Trump calling Kim "little rocket man" and Kim labeling Trump a "dotard" and "old lunatic." Trump also engaged in a bitter Twitter battle with French President Emmanuel Macron and Canadian Prime Minister Justin Trudeau about tariffs and trade that prompted the American president to leave the G7 summit early in 2018 (Liptak, Kosinski, and Diamond, 2018).

### The Political Twittersverse

The emergence of the political "Twittersverse" has coincided with the rise of social media over the past decade. Social media are distinct from other digital channels in that they host interactions among users who set up personal profiles and communicate with others in their networks (Carr and Hayes, 2015). Participants use social media to create and distribute content, consume and interact with material posted by their connections, and share their views publicly (Ellison and Boyd, 2013). Social-networking sites can help people to maintain and develop social relationships, engage in social surveillance and voyeurism, and pursue self-promotion (Alhabash and Ma, 2017). Users often employ social media to seek out and follow like-minded people and groups which promotes social bonding, reinforces personal and political identities, and provides digital companionship (Jung and Sundar, 2016; Joinson, 2008). These characteristics are highly conducive to the adaptation of social media—especially Twitter—for political use. Donald Trump's Twitter followers identify with him on a personal level, which encourages their blanket acceptance of his policies (Owen, 2018).

**Origins of Social Media** The current era of networked communication originated with the invention of the World Wide Web in 1991, and the development of Weblogs, list-serves, and e-mail that supported online communities. The first social-media site, Six Degrees, was developed in 1997 and disappeared in 2000 as there were too few users to sustain it. Niche sites catering to identity groups, such as Asian Avenue, friendship circles, including Friendster and MySpace, professional contacts, like LinkedIn, and public-policy advocates, such as MoveOn, also emerged in the late 1990s and early 2000s. The advent of Web 2.0 in the mid-2000s, with its emphasis on participatory, user-centric, collaborative platforms, coincided with the development of enduring social-media platforms, including Facebook (2004) and YouTube (2005) (van Dijck, 2013; Edosomwan et al., 2011), which have become staples of political campaigning, organizing, and governing.

When Twitter was founded in 2006, it was envisioned as a microblogging site where groups of friends could send short messages (tweets) to their friends about what was happening in their lives in real time in a manner akin to texting. Users could also upload photos, GIFs, and short videos to the site. Twitter initially imposed a 140-character constraint on tweets which was the limit that mobile carriers placed on SMS text messages. The limit was increased to 280 characters in 2017 as the popularity of the platform peaked and wireless carrier restrictions



on the amount of content users could send were no longer relevant. Users easily circumvent the character limit by posting multiple tweets in sequence, a practice used frequently by Donald Trump. Twitter's user base quickly sought to expand the functionality of the platform by including the @ symbol before a username to identify other users and adding #hashtags to mark content, making it easier to follow topics and themes. While the actual number of Twitter followers is difficult to verify as many accounts are dormant or controlled by bot software, it is estimated that there were 335 million active monthly users worldwide in 2018 (Statista, 2018). Still, it took Twitter twelve years to turn a profit for the first time when it increased attention to ad sales (Tsukayama, 2018).

**Twitter in the Age of Trump** Standard communication practices for American politicians have been upended by Donald Trump's use of Twitter. Trump was on Twitter seven years prior to his quest for the presidency. His signature Twitter style was established during his days on the reality TV program *The Apprentice*, and it has changed little since he launched his political career. Trump's Twitter pronouncements are superficial, which is well-suited to a medium with a word limit on posts. His messages are expressed conversationally using an accessible, informal, fourth-grade vocabulary. Trump's tweets have the tone of a pitchman who is trying to sell a bill of goods (Grieve and Clarke, 2017; Clarke and Grieve, 2017). They are often conflicting, confusing, and unclear, which allows users to interpret them based on their own preconceptions. His posts are repetitious and have a similar cadence which fosters a sense of believability and familiarity among his loyalists (Graham, 2018).

## **Social-media users often employ social media to seek out and follow like-minded people and groups which promotes social bonding, reinforces personal and political identities, and provides digital companionship**

Trump has bragged that his control over Twitter paved his path to the White House (Tatum, 2017). As a presidential candidate, Trump effectively engaged Twitter to publicize his thoughts, attack his long list of enemies, and hijack political discourse. His supporters became ardent followers of his Twitter messages during the campaign. Trump's campaign social-media style contrasted with Hillary Clinton's more controlled approach, which had been the pre-Trump norm (Enli, 2017). Clinton's social-media posts were measured in tone, rarely made personal attacks, and provided reasons and facts supporting her issue positions. In contrast, Trump made broad, general declarations that lacked evidence (Stromer-Galley, 2016) and claimed credit for the accomplishments of others (Tsur et al., 2016).

Since the election, Twitter has become Trump's connection to the world outside the White House. He engages in "rule by tweet," as his Twitter feed substitutes for regular presidential press conferences and weekly radio addresses that were the norm for his predecessors when making major policy announcements. Trump regularly produces nasty and outlandish tweets to ensure that he remains at the center of attention, even as political and natural disasters move the spotlight. Twitter changes the life cycle of news as developing reports can be readily overtaken by a new story generated by a provocative tweet. A digital communications strategist for the Trump campaign named Cassidy asserted that this strategy has been intentional: "Trump's goal from the beginning of his candidacy has been to set the agenda of the media. His strategy is to keep things moving so fast, to talk so loudly—literally and metaphorically—that

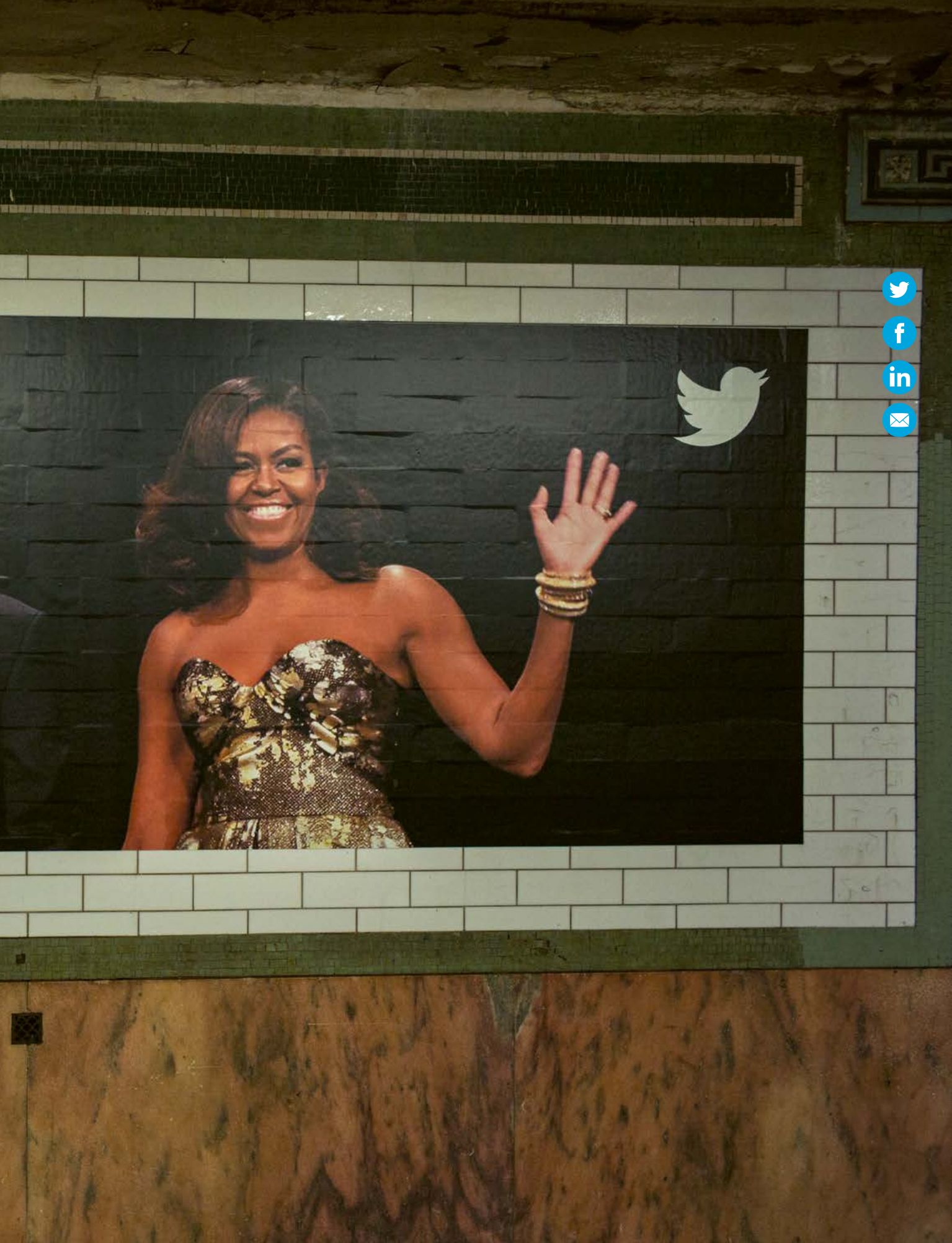
A large mural of Barack Obama in a dark suit and tie, set against a background of red and black geometric shapes. A hand is reaching up from the mural towards the Twitter logo. The mural is mounted on a wall of white subway tiles.

**The political Twittersverse  
has radically changed how  
politicians, citizens and the  
press transmit information**

A Twitter advertisement at a New York subway station shows ex-president of the United States Barack Obama with his wife, Michelle Obama







- [Twitter](#)
- [Facebook](#)
- [LinkedIn](#)
- [Email](#)



the media, and the people, can't keep up" (Woolley and Guilbeault, 2017: 4). Trump's Twitter tirades often sideline public discussions of important policy issues, such as immigration and health care, distract from embarrassing personal scandals, and attempt to camouflage the mishaps of his administration.

**Amplification, Incivility, and Polarization** The power of social media to influence politics is enhanced due to their ability to amplify messages quickly through diverse media platforms. Social media have become a steady source of political content for news outlets with large audiences, especially cable news. Trump supporters regularly tune in to Fox News to cheer his latest Twitter exploits. Talking heads on media that view Trump more negatively, like CNN and MSNBC, spend countless hours attempting to interpret his tweets as the messages are displayed prominently on screen. Supporters who interact with, annotate, and forward his messages lend greater credence to Trump's missives within their networks. Trump's individual messages are retweeted 20,000 times on average; some of his anti-media screeds have been retweeted as many as 50,000 times (Wallsten, 2018). The perception that Trump is powerful is enhanced simply by virtue of the amount of attention he draws. Politicians can use the amplification power of social media to galvanize public opinion and to call people to action. These benefits of Twitter amplification have been shown to empower men substantially more than women. Male political journalists have a greater audience on Twitter than their female peers, and they are more likely to spread the political messages of men than women (Usher, Holcomb, and Littman, 2018).

## **The power of social media to influence politics is enhanced due to their ability to amplify messages quickly through diverse media platforms. Social media have become a steady source of political content for news outlets with large audiences, especially cable news**

Social media host discourse that is increasingly incivil and politically polarizing. Offhanded remarks are now released into the public domain where they can have widespread political consequences (van Dijck, 2013). Trump's aggressive Twitter pronouncements are devoid of filtering that conforms to cultural codes of decency. He makes expansive use of adjectives, typically to describe himself in positive terms and to denigrate others. He constantly refers to himself as "beautiful," "great," "smartest," "most successful," and "having the biggest brain." In contrast, he refers to those who fall out of favor as "lying Hillary Clinton," "untruthful slime ball James Comey" (a former FBI director), and "crazy Mika" (Brzezinski, a political talk-show host). As of June 2018, Trump had used the terms "loser" 246 times, "dummy" 232 times, and "stupid" 192 times during his presidency to describe people he dislikes (Armstrong, 2018).

Politicians have engaged in partisan Twitter wars that have further divided Republicans and Democrats, conservatives and liberals. On a day when the *New York Times* revealed that the Trump family had dodged millions of dollars in taxes and provided Donald Trump with far more financial support than he has claimed, Trump sought to shift the press agenda by tweeting his anger at "the vicious and despicable way Democrats are treating [Supreme Court nominee] Brett Kavanaugh!" Research conducted at the University of Pennsylvania found that the political use of Twitter by the general public is concentrated among a small subset of the public representing polar ideological extremes who identify as either "very conserv-

ative” or “very liberal.” Moderate political voices are rarely represented on the platform (Preotiuc-Pietro et al., 2017).

The Twittersverse is highly prone to deception. Twitter and other social media have contributed to—even fostered—the proliferation of false information and hoaxes where stories are entirely fabricated. False facts spread fast through social media. They can make their way onto legitimate news platforms and are difficult to rebut as the public has a hard time determining fact from fiction. Many people have inadvertently passed on “fake news” through their social-media feeds. In fact, an MIT study found that people are more likely to pass on false stories through their networks because they are often novel and generate emotional responses in readers (Vosoughi, Roy, and Aral, 2018). President Barack Obama has referred to the confusion caused by conspiracy theories and misleading information during the 2016 election as a “dust cloud of nonsense” (Heath, 2016).

Misinformation is often targeted at ideological audiences, which contributes the rise in political polarization. A *BuzzFeed News* analysis found that three prominent right-wing Facebook pages published misinformation 38% of the time and three major left-wing pages posted false facts 20% of the time (Silverman et al., 2016). The situation is even more severe for Twitter, where people can be completely anonymous and millions of automated bots and fake accounts have flooded the network with tweets and retweets. These bots have quickly outrun the spam detectors that Twitter has installed (Manjoo, 2017).

## **Trump supporters regularly tune in to Fox News to cheer his latest Twitter exploits. Talking heads on media that view Trump more negatively, like CNN and MSNBC, spend countless hours attempting to interpret his tweets as the messages are displayed prominently on screen**

The dissemination of false information through Twitter is especially momentous amidst the uncertainty of an unfolding crisis where lies can spread much faster than the truth (Vosoughi, Roy, and Aral, 2018). Misinformation and hoaxes circulated widely as the shooting of seventeen students at Marjory Stoneman Douglas High School in Parkland Florida in February of 2018 was unfolding. Conspiracy theories about the identity of the shooter pointed in the wrong direction. False photos of the suspect and victims were circulated. Tweets from a reporter with the *Miami Herald* were doctored and retweeted to make it appear as if she had asked students for photos and videos of dead bodies. Within an hour of the shooting, Twitter accounts populated by Russian bots circulated hundreds of posts about the hot-button issue of gun control designed to generate political divisiveness (Frenkel and Wakabayashi, 2018). In the weeks after the shooting, as Parkland students became activists for stronger gun control measures, conspiracy theories proliferated. A widespread rumor asserted that the students were “crisis actors” who had no affiliation with the school. A doctored conspiracy video attacking the students was posted on YouTube and became a top trending clip on the site (Arkin and Popken, 2018).

### **The Emergence of News Deserts**

The consequences of the rise of social media and the spread of false information have been elevated by the disappearance of trusted local news organizations from the media landscape.





The proliferation of “news deserts”—communities where there are no responsible local news organizations to provide information to residents and to counter false stories—has meant that misinformation is often taken for fact and spread virally through people’s social networks unchecked (Bucay et al., 2017). Facebook and Twitter have been reluctant to deal effectively with the flow of misinformation through their platforms and have even refused to remove demonstrated false stories from their sites (Hautala, 2018). In 2018, Facebook, Twitter, Apple, and YouTube permanently banned alt-right radio host Alex Jones and his site, *InfoWars*, after failing to take any disciplinary action for years against his prolific spread of abusive conspiracy theories. But this move by big media companies was an exception. These circumstances have raised the potential for misinformation to unsettle the political system.

## **The proliferation of “news deserts”—communities where there are no responsible local news organizations to provide information to residents and to counter false stories—has meant that misinformation is often taken for fact and spread virally through people’s social networks**

Local news historically has been a staple of Americans’ media diets. Less than a decade ago, local newspapers were responsible for approximately 85% of fact-based and investigative news (Jones, 2011). However, their importance for informing and shaping opinions of people in small towns and suburban communities has often been underestimated. Community newspapers have been far more consequential to their millions of readers than large newspapers of record, such as *The New York Times* and *The Washington Post*, whose stories are amplified on twenty-four-hour cable news programs. People tend to trust their local news outlets, and to have faith that the journalists—who are their neighbors—will report stories accurately. Citizens have relied heavily on local news outlets to keep them informed about current happenings and issues that are directly relevant to their daily lives. Local news stories influence the opinions of residents and routinely impact the policy decisions made by community leaders. Importantly, audiences rely on local journalists to provide them with the facts and to act as a check on misinformation that might be disseminated by outside sources, especially as they greatly distrust national news (Abernathy, 2016).

Trump’s tweets have become more relevant to his base in an era when local news sources have been disappearing from the media landscape. His tweets can substitute for news among people living in news deserts. The influence of his social-media messaging is enhanced in places where access to local media that would check his facts, provide context for his posts, and offer alternative interpretations is lacking. The dearth of robust local news outlets is especially pronounced in rural, predominantly white areas of the country where Donald Trump’s political base is ensconced (Lawless and Hayes, 2018; Center for Innovation and Sustainability in Local Media, 2017). With the counterbalance of trusted local news sources, Trump’s attacks on the mainstream media resonate strongly, and the public’s political perspective is heavily influenced by relentless partisan social-media messages that masquerade as news in their feed (Musgrave and Nussbaum, 2018).

The concentration of media ownership in the hands of large corporations has further undermined truly local news. As independent media organizations have disappeared, they have been replaced in an increasing number of markets by platforms owned by news conglomerates. Sinclair Broadcast Group is the largest television news conglomerate in the United States and





A New Delhi newspaper seller holds a daily whose back page advertisement is a Whatsapp warning about fake news of child kidnapping in India that went viral on social media and led to lynchings throughout the country in summer, 2018





has bought out local stations across the country, including in news deserts. The company has strong ties to the Trump administration and has pushed its reporters to give stories a more conservative slant. An Emory University study revealed that Sinclair's TV news stations have shifted coverage away from local news to focus on national stories, and that coverage has a decidedly right-wing ideological perspective (Martin and McCrain, 2018). On a single day, Sinclair compelled their local news anchors to give the same speech warning about the dangers of "fake news," and stating that they were committed to fair reporting. Many anchors were uncomfortable making the speech, which they described as a "forced read." A video of anchors across the country reciting the script was widely reported in the mainstream press and went viral on social media (Fortin and Bromwich, 2018).

## The Future

The digital revolution has unfolded more rapidly and has had broader, deeper, and more transformative repercussions on politics and news than any prior transition in communication technology, including the advent of television. Over the past decade, the rise in social media as a political tool has fundamentally changed the relationships between politicians, the press, and the public. The interjection of Donald Trump into the political media mix has hastened the evolution of the media system in some unanticipated directions. As one scholar has noted: "The Internet reacted and adapted to the introduction of the Trump campaign like an ecosystem welcoming a new and foreign species. His candidacy triggered new strategies and promoted established Internet forces" (Persily, 2017).

The political media ecology continues to evolve. Politicians are constantly seeking alternatives to social media as a dominant form of communicating to the public. Candidates in the 2018 midterm elections turned to text messages as a campaign tool that is supplanting phone-banks and door-to-door canvassing as a way of reaching voters. New developments in software, such as Hustle, Relay, and RumbleUp, have made it possible to send thousands of texts per hour without violating federal laws that prohibit robo-texting—sending messages in bulk. Texts are used to register voters, organize campaign events, fundraise, and advertise. The text messages are sent by volunteers who then answer responses from recipients. The strategy is aimed especially at reaching voters in rural areas and young people from whom texting is the preferred method of digital communication (Ingram, 2018). Much like Trump's Twitter feed, texting gives the perception that politicians are reaching out personally to their constituents. The tactic also allows politicians to distance themselves from big media companies, like Facebook and Google, and the accompanying concerns that personal data will be shared without consent.

**The digital revolution has unfolded more rapidly and has had broader, deeper, and more transformative repercussions on politics and news than any prior transition in communication technology, including the advent of television**

Great uncertainty surrounds the future of political communication. The foregoing discussion has highlighted some bleak trends in the present state of political communication. Political polarization has rendered reasoned judgment and compromise obsolete. The rampant spread

of misinformation impedes responsible decision-making. The possibility for political leaders to negatively exploit the power of social media has been realized.

At the same time, pendulums do shift, and there are positive characteristics of the new media era that may prevail. Digital media have vastly increased the potential for political information to reach even the most disinterested citizens. Attention to the 2018 midterm elections was inordinately high, and the ability for citizens to express themselves openly through social media has contributed to this engagement. Issues and events that might be outside the purview of mainstream journalists can be brought to prominence by ordinary citizens. Social media has kept the #MeToo movement alive as women continue to tell their stories and form online communities. Further, there is evidence of a resurgence in investigative journalism that is fueled, in part, by access to vast digital resources available for researching stories, including government archives and big data analysis. These trends offer a spark of hope for the future of political media.

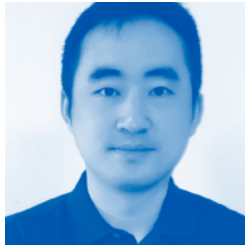


## Select Bibliography

- Abernathy, Penelope Muse. 2016. *The Rise of a New Media Baron and the Emerging Threat of News Deserts*. Research Report. Chapel Hill, NC: The Center for Innovation and Sustainability in Local Media, School of Media and Journalism, University of North Carolina at Chapel Hill.
- Alhabash, Saleem, and Ma, Mangyan. 2017. "A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students?" *Social Media and Society*, January-March: 1–13.
- Arkin and Popken, 2018. "How the Internet's conspiracy theorists turned Parkland students into 'crisis actors.'" *NBC News*, February 21. <https://www.nbcnews.com/news/us-news/how-internet-s-conspiracy-theorists-turned-parkland-students-crisis-actors-n849921>.
- Armstrong, Martin. 2018. "Trump's favorite Twitter insults." *Statista*, June 28. <https://www.statista.com/chart/14474/trump-favorite-twitter-insults/>.
- Bimber, Bruce. 2014. "Digital media in the Obama campaigns of 2008 and 2012: Adaptation to the personalized political communication environment." *Journal of Information Technology & Politics* 11(2): 130–150.
- Bucay, Yemile, Elliott, Vitoria, Kamin, Jennie, and Park, Andrea. 2017. "America's growing news deserts." *Columbia Journalism Review*, spring. [https://www.cjr.org/local\\_news/american-news-deserts-donuts-local.php](https://www.cjr.org/local_news/american-news-deserts-donuts-local.php).
- Carr, Caleb T., and Hayes, Rebecca. 2015. "Social media: Defining, developing, and divining." *Atlantic Journal of Communication* 23(1): 46–65.
- Center for Innovation and Sustainability in Local Media. 2017. *Thwarting the Emergence of News Deserts*. Chapel Hill, NC: University of North Carolina School of Media and Journalism. <https://www.cislm.org/wp-content/uploads/2017/04/Symposium-Leave-Behind-0404finalweb.pdf>.
- Clarke, Isabelle, and Grieve, Jack. 2017. "Dimensions of abusive language on Twitter." In *Proceedings of the First Workshop on Abusive Language Online*. Vancouver, Canada: The Association of Computational Linguistics, 1–10. <http://www.aclweb.org/anthology/W17-3001>.
- Davis, Richard, and Owen, Diana. 1998. *New Media in American Politics*. New York: Oxford University Press.
- Edosomwan, Simeon, Prakasan, Sitalaskshmi Kalangot, Kouame, Doriane, Watson, Jonelle, and Seymour, Tom. 2011. "The history of social media and its impact on business." *The Journal of Applied Management and Entrepreneurship* 16(3). [https://www.researchgate.net/profile/Simeon\\_Edosomwan/publication/303216233\\_The\\_history\\_of\\_social\\_media\\_and\\_its\\_impact\\_on\\_business/links/57fe90ef08ae56fae5f23f1d/The-history-of-social-media-and-its-impact-on-business.pdf](https://www.researchgate.net/profile/Simeon_Edosomwan/publication/303216233_The_history_of_social_media_and_its_impact_on_business/links/57fe90ef08ae56fae5f23f1d/The-history-of-social-media-and-its-impact-on-business.pdf).
- Ellison, Nicole B., and Boyd, Danah. 2013. "Sociality through social network sites." In *The Oxford Handbook of Internet Studies*, William H. Dutton (ed.). Oxford, UK: Oxford University Press, 151–172.
- Enli, Gunn. 2017. "Twitter as arena for the authentic outsider: Exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election." *European Journal of Communication* 32(1): 50–61.
- Fortin, Jacey, and Bromwich, Jonah Engel. 2018. "Sinclair made dozens of local news anchors recite the same script." *The New York Times*, April 2. <https://www.nytimes.com/2018/04/02/business/media/sinclair-news-anchors-script.html>.
- Frenkel, Sheera, and Wakabayashi, Daisuke. 2018. "After Florida school shooting, Russian 'bot' army pounced." *The New York Times*, February 2. <https://www.nytimes.com/2018/02/19/technology/russian-bots-school-shooting.html>.
- Gottfried, Jeffrey, and Shearer, Elisa. 2017. "Americans' online news use is closing in on TV news use." *Fact Tank*. Washington, DC: Pew Research Center. <http://www.pewresearch.org/fact-tank/2017/09/07/americans-online-news-use-vs-tv-news-use/>.
- Gottfried, Jeffrey, and Barthel, Michael. 2018. "Almost seven-in-ten Americans have news fatigue, more among Republicans." Washington, DC: Pew Research Center. <http://www.pewresearch.org/fact-tank/2018/06/05/almost-seven-in-ten-americans-have-news-fatigue-more-among-republicans/>.
- Graham, David A. 2018. "The gap between Trump's tweets and reality." *The Atlantic*, April 2. <https://www.theatlantic.com/politics/archive/2018/04/the-world-according-to-trump/557033/>.
- Grieve, Jack, and Clarke, Isabelle. 2017. "Stylistic variation in the @realDonaldTrump Twitter account and the stylistic typicality of the pled tweet." Research Report. University of Birmingham, UK. <http://rpubs.com/jwgrieve/340342>.
- Guo, Lei. 2015. "Exploring the link between community radio and the community: A study of audience participation in alternative media practices." *Communication, Culture, and Critique* 10(1). <https://onlinelibrary.wiley.com/doi/full/10.1111/cccr.12141>.
- Hautala, Laura. 2018. "Facebook's fight against fake news remains a bit of a mystery." *c/net*, July 24. <https://www.cnet.com/news/facebook-says-misinformation-is-a-problem-but-wont-say-how-big/>.
- Heath, Alex. 2016. "OBAMA: Fake news on Facebook is creating a 'dust cloud of nonsense.'" *Business Insider*, November 7. <https://www.businessinsider.com/obama-fake-news-facebook-creates-dust-cloud-of-nonsense-2016-11>.
- Hindman, Matthew. 2018. *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy*. Princeton, NJ: Princeton University Press.
- Ingram, David. 2018. "Pocket partisans: Campaigns use text messages to add a personal touch." *NBC News*, October 3. <https://www.nbcnews.com/tech/tech-news/sofas-ipads-campaign-workers-use-text-messages-reach-midterm-voters-n915786>.
- Joinson, Adam N. 2008. "'Looking at,' 'looking up' or 'keeping up with' people? Motives and uses of Facebook." *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*. Florence, Italy, April 5–10: 1027–1036.
- Jones, Alex S. 2011. *Losing the News: The Future of the News that Feeds Democracy*. New York: Oxford University Press.
- Jost, John T., Barberá, Pablo, Bonneau, Richard, Langer, Melanie, Metzger, Megan, Nagler, Jonathan, Sterling, Joanna, and Tucker, Joshua A. 2018. "How social media facilitates political protest: Information, motivation, and social networks." *Political Psychology* 39(51). <https://onlinelibrary.wiley.com/doi/full/10.1111/pops.12478>.
- Jung, Eun Hwa, Sundar, Shyam S. 2016. "Senior citizens on Facebook: How do they interact and why?" *Computers in Human Behavior* 61: 27–35.
- Lawless, Jennifer, and Hayes, Danny. 2018. "As local news goes, so goes citizen engagement: Media, knowledge, and participation in U.S. House elections." *Journal of Politics* 77(2): 447–462.
- Liptak, Kevin, Kosinski, Michelle, and Diamond, Jeremy. 2018. "Trump to skip climate portion of G7 after Twitter spat with Macron and Trudeau." *CNN* June 8. <https://www.cnn.com/2018/06/07/politics/trump-g7-canada/index.html>.
- Manjoo, Farhad. 2017. "How Twitter is being gamed to feed misinformation." *The New York Times*, May 31. <https://www.nytimes.com/2017/05/31/technology/how-twitter-is-being-gamed-to-feed-misinformation.html>.
- Martin, Gregory J., and McCrain, Josh. 2018. "Local news and national politics." Research Paper. Atlanta, Georgia: Emory University. <http://joshuamccrain.com/localnews.pdf>.
- Matsa, Katerina Eva. 2018. "Fewer Americans rely on TV news: What type they watch varies by who they are." Washington, DC: Pew Research Center, January 5. <http://www.pewresearch.org/fact-tank/2018/01/05/fewer-americans-rely-on-tv-news-what-type-they-watch-varies-by-who-they-are/>.



- Musgrave, Shawn, and Nussbaum, Matthew. 2018. "Trump thrives in areas that lack traditional news outlets." *Politico*, March 8. <https://www.politico.com/story/2018/04/08/news-subscriptions-decline-donald-trump-voters-505605>.
- Owen, Diana. 2017a. "The new media's role in politics." In *The Age of Perplexity: Rethinking the World We Know*, Fernando Gutiérrez Junquera (ed.). London: Penguin Random House.
- Owen, Diana. 2017b. "Tipping the balance of power in elections? Voters' engagement in the digital campaign." In *The Internet and the 2016 Presidential Campaign*, Terri Towner and Jody Baumgartner (eds.). New York: Lexington Books, 151–177.
- Owen, Diana. 2018. "Trump supporters' use of social media and political engagement." Paper prepared for presentation at the Annual Meeting of the American Political Science Association, August 30–September 2.
- Persily, Nathaniel. 2017. "Can democracy survive the Internet?" *Journal of Democracy* 28(2): 63–76.
- Pew Research Center. 2018. "Social media fact sheet: Social media use over time." February 5, 2018. Washington, DC: Pew Research Center. <http://www.pewinternet.org/fact-sheet/social-media/>.
- Preotiuc-Pietro, Daniel, Hopkins, Daniel J., Liu, Ye, and Unger, Lyle. 2017. "Beyond binary labels: Political ideology prediction of Twitter users." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Vancouver, Canada: Association for Computational Linguistics, 729–740. <http://aclweb.org/anthology/P/P17/P17-1068.pdf>.
- Rohlinger, Deana A., and Bunnage, Leslie. 2017. "Did the Tea Party movement fuel the Trump train? The role of social media in activist persistence and political change in the 21st century." *Social Media + Society* 3(2). <http://journals.sagepub.com/doi/full/10.1177/2056305117706786>.
- Sayej, Nadja. 2017. "Alyssa Milano on the #MeToo movement: 'We're not going to stand for it any more.'" *The Guardian*, December 1. <https://www.theguardian.com/culture/2017/dec/01/alyssa-milano-mee-too-sexual-harassment-abuse>.
- Shearer, Elisa, and Matsa, Katerina Eva. 2018. *News Across Social Media Platforms 2018*. Research Report. Washington, DC: Pew Research Center, September 10. <http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>.
- Silverman, Craig, Hall, Ellie, Strapagiel, Lauren, Singer-Vine, Jeremy, and Shaban, Hamza. 2016. "Hyperpartisan Facebook pages are publishing false and misleading information at an alarming rate." *BuzzFeed News*, October 20. <https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis#.wDLRgawza>.
- Smith, Aaron, and Anderson, Monica. 2018. *Social Media Use in 2018*. Research Report. Washington, DC: Pew Research Center. <http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>.
- Statista. 2018. "Number of monthly active Twitter users worldwide from 1st quarter 2018 (in millions)." <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- Stromer-Galley, Jennifer. 2016. *Presidential Campaigning in the Internet Age*. New York: Oxford University Press.
- Tatum, Sophie. 2017. "Trump on his tweets: 'You have to keep people interested.'" *CNN Politics*, October 21. <https://www.cnn.com/2017/10/20/politics/donald-trump-fox-business-interview-twitter/index.html>.
- Tsukayama, Hayley. 2018. "Why Twitter is now profitable for the first time." *The Washington Post*, February 8. [https://www.washingtonpost.com/news/the-switch/wp/2018/02/08/why-twitter-is-now-profitable-for-the-first-time-ever/?utm\\_term=.d22610b684af](https://www.washingtonpost.com/news/the-switch/wp/2018/02/08/why-twitter-is-now-profitable-for-the-first-time-ever/?utm_term=.d22610b684af).
- Tsur, Oren, Ognyanova, Katherine, and Lazer, David. 2016. "The data behind Trump's Twitter takeover." *Politico Magazine* April 29. <https://www.politico.com/magazine/story/2016/04/donald-trump-2016-twitter-takeover-213861>.
- Usher, Nikki, Holcomb, Jesse, and Littman, Justin. 2018. "Twitter makes it worse: Political journalists, gendered echo chambers, and the amplification of gender bias." *International Journal of Press/Politics* 1–21. <http://journals.sagepub.com/doi/10.1177/1940161218781254>.
- Van Dijck, José. 2013. *The Culture of Connectivity: A Critical History of Social Media*. New York: Oxford University Press.
- Vosoughi, Soroush, Roy, Deb, and Aral, Sinan. 2018. "The spread of true and false news online." *Science* 359(6380): 1146–1151.
- Wallsten, Kevin. 2010. "'Yes We Can': How online viewership, blog discussion, campaign statements, and mainstream media coverage produced a viral video phenomenon." *Journal of Information Technology and Politics* 2(3): 163–181.
- Wallsten, Scott. 2018. "Do likes of Trump's tweets predict his popularity?" *Technology Policy Institute Blog*, April 17. <https://techpolicyinstitute.org/2018/04/17/do-likes-of-trumps-tweets-predict-his-popularity/>.
- Woolley, Samuel C., and Guilbeault, Douglas R. 2017. "Computational propaganda in the United States of America: Manufacturing consensus online." Working Paper No. 2017.5. Oxford Internet Institute: Computational Propaganda Research Project.



**Yang Xu**  
Hong Kong Polytechnic  
University

Yang Xu is an Assistant Professor in the Department of Land Surveying and Geo-Informatics (LSGI) at the Hong Kong Polytechnic University. His research lies at the intersection of geographic information science, transportation, and urban informatics. By harnessing the power of big data, Xu's research answers important questions of individual travel behavior, collective human mobility patterns, and various aspects of urban and social dynamics. In the past couple of years, he has coauthored more than twenty publications in prestigious journals and books. He also served as the principal investigator of several projects funded by the Hong Kong Research Grant Council and National Natural Science Foundation of China. Xu earned his BSc and MSc at Wuhan University, and later his PhD at the University of Tennessee, Knoxville. Before joining LSGI, Xu worked as a joint postdoctoral associate at the MIT Senseable City Lab and the Singapore-MIT Alliance for Research and Technology (SMART).



**Carlo Ratti**  
MIT-Massachusetts Institute  
of Technology

An architect and engineer by training, Professor Carlo Ratti teaches at MIT, where he directs the Senseable City Laboratory, and is a founding partner of the international design and innovation practice Carlo Ratti Associati. A leading voice in the debate on new technologies' impact on urban life, his work has been exhibited in several venues worldwide, including the Venice Biennale, New York's MoMA, London's Science Museum, and Barcelona's Design Museum. Two of his projects—the Digital Water Pavilion and the Copenhagen Wheel—were hailed by *Time* magazine as "Best Inventions of the Year." He has been included in *Wired* magazine's "Smart List: 50 people who will change the world." He is currently serving as co-chair of the World Economic Forum's Global Future Council on Cities and Urbanization, and as special advisor on Urban Innovation to the European Commission. For further information visit [www.carloratti.com](http://www.carloratti.com) and [senseable.mit.edu](http://senseable.mit.edu).

Recommended book: *The City of Tomorrow*, Carlo Ratti and Matthew Claudel, Yale University Press, 2016.

**Considering the recent increase of socioeconomic segregation in cities worldwide, scholars of urban technology Carlo Ratti and Yang Xu put forward a new metric with which to study and counter such a phenomenon. While past research has focused on residential segregation, this article argues that segregation needs to be tracked more dynamically: across the urban environment and not just at home; through time and not just space; and by monitoring its presence in social space and not just physical space. These methods require more dynamic data as well: Ratti and Xu argue for the greater cost-effectiveness and possibilities presented by mobile-phone data. They conclude by speculating on some design actions cities could take to encourage greater mixing between different groups.**



The past half century has witnessed an accelerated process of global urbanization, bringing profound changes to the spatial distribution of the world's population and the underlying human activities. In 2007, the world's urban population surpassed, for the first time in history, the global rural population (United Nations, 2014). It is projected that, by 2050, more than two thirds of the world population will live in urban areas, a good fraction of them in so-called "megacities." Such agglomerations will bring numerous social and economic benefits to urban dwellers. However, rapid urbanization might also cause or exacerbate certain issues. Social segregation, a long-standing challenge for cities, remains to be a headache for urban governors and policy makers.

Urban segregation, by its classic definition, refers to the physical separation or uneven distribution of social groups in cities. Social groups, in this context, are usually measured using their social, economic, and/or demographic characteristics. The variations in how social groups are distinguished result in different perspectives of urban segregation, covering the racial, ethnic, income, and other aspects of social mixings in cities.

Why tackling segregation is becoming so pressing? One crucial reason is that segregation is usually tied to inequality, which marks the unequal access of different social groups to certain resources or benefits. The physical sorting of social groups in cities, which is just an appearance of urban segregation, could lead to many fundamental issues that are detrimental to the well-being of societies. Racial segregation in the United States, for example, has limited access to jobs, education, and public services for people who live in high-poverty neighborhoods (Williams and Collins, 2001). These are not problems just for the Americans. Over the past decade, there has been an increase in socioeconomic segregations in many European cities (Musterd et al., 2017), causing issues of rising poverty, crime, and even terrorism. In Asia, many megacities and urban agglomerations are expected to form in the next twenty to thirty years. Part of the low-income neighborhoods or areas that used to be on the outskirts of the cities will become "urban villages," separating the disadvantaged groups from the others. What could make the matter worse is that gentrification processes might drive these people out of their habitats. Rising housing prices might force them to move further away from the city centers, resulting in longer commuting distances and decreased accessibility to health facilities and other urban amenities.

**In 2007, the world's urban population surpassed, for the first time in history, the global rural population. It is projected that, by 2050, more than two thirds of the world population will live in urban areas, a good fraction of them in so-called "megacities"**

Being aware of the problem does not guarantee a solution. To tackle urban segregation, the initial and important step is to understand to what extent different social groups are separated in cities. In the past few decades, considerable efforts have been devoted in cities to investigate *residential segregation* (Massey and Denton, 1988). There are two main reasons to focus on this. First, residential location and its surroundings—as one's primary activity territory—have a profound influence on people's daily social interactions. A high level of residential segregation signifies social stratification and serves as a driving force of many societal issues. The second reason is related to the limitation of traditional data-collection



techniques. Population census in many countries is conducted every five or ten years. Such data capture static snapshots of population footprint and demographic information. They are suitable for measuring residential segregation but are unable to deliver a timely view of socioeconomic configurations in cities.

Would it be useful to measure urban segregations beyond residence? The answer is obvious. Social segregations do not only take place where people live, but also where people work, study, and entertain themselves (Hellerstein and Neumark, 2008). Different types of places could play different roles in dampening or strengthening a city's social integration. An improved understanding of social dynamics requires observing human interactions at workplaces, schools, shopping malls, and all kinds of public spaces. However, a comprehensive view of urban segregation cannot be achieved by only focusing on the spatial dimension. How do the social mixings of a city change over time? Residential segregation already gives us a night-time story, but what about the counterpart during daytime? Such a temporal perspective is desired simply because cities are rather dynamic. As people get up in the morning and start to move, the changes in urban mobility would reshape the socioeconomic configurations of a city. The impacts of human mobility on segregation as well as its policy implications remain to be better understood.

Social segregations do not only exist in physical space. In the past two decades, Internet and telecommunication technologies have permeated almost every aspect of human life, moving the world toward a place with ubiquitous connectivity. The technological advancements have created a digital world in which new forms of social communications are emerging. How will these changes affect the social structures of cities? Is there a stratification in the social world, and whether it is simply an image of how people interact in physical space? Answering these questions could deliver a more holistic view of social mixings in cities. Unfortunately, the data we used to rely on, such as census and travel surveys, are unable to capture human social interactions in the "social space." There is a need to redefine how we address urban segregations in today's world, and it is not only about new data, but also new metrics, and new strategies for tackling this enduring urban issue.

In this article, we first review some of the leading research in this field. We then provide some thoughts on how to develop a new metrics for quantifying urban segregation. Finally, we discuss how the metric can possibly inspire new design principles for better integrating socioeconomic classes in cities.

### Measuring Residential Segregation: Some Existing Efforts

Understanding the spatial separation or concentration of different population groups in physical space has long been a research interest in sociology, economics, geography, urban planning, and other fields. For decades, studies have primarily focused on investigating segregation at places of residence (Massey and Denton, 1987; Reardon and O'Sullivan, 2004), and a series of methods have been developed to quantify the uneven distribution of social groups in physical space. The index of dissimilarity (DI) is one of the most commonly used methods in segregation studies (Duncan and Duncan, 1955). The index, which was often used in two-group cases (for example, segregation between black and white), is interpreted as the proportion of minority members that would have to change their area of residency to achieve a perfect social integration (that is, an even distribution in all areas), normalized by the proportion of minority members that would have to move under conditions of maximum segregation (that is, no areas are shared by the two groups). Since it was developed, the dissimilarity in-





A man lays trouser to dry on a tin roof in an outdoor laundry known as Dhobi Ghat, near a neighborhood of luxury residential skyscrapers in Mumbai, India. It is estimated that Mumbai's population will surpass 65 million inhabitants by 2100





dex was employed in many studies due to its simplicity in representing segregation. Despite the heavy use of DI, the definitions of residential segregation vary among researchers and many other indices have been developed to quantify segregation from different perspectives. This has triggered extensive discussions on the redundancy of segregation indices as well as which ones should be mainly adopted (Cortese et al., 1976; Taeuber and Taeuber, 1976). By examining twenty segregation indices using factor analysis, Massey and Denton (1988) concluded that these indices mainly explain five distinct dimensions of residential segregation, which are *evenness*, *exposure*, *concentration*, *centralization*, and *clustering*. Evenness measures the imbalance of two or more population groups in different neighborhoods, while exposure quantifies the interaction potential among social groups. These two dimensions have attracted more attention in empirical studies than the other three.

## **The index of dissimilarity (DI), one of the most commonly used methods in segregation studies, is interpreted as the proportion of minority members that would have to change their area of residency to achieve a perfect social integration**

Since Massey and Denton's reflections on the five principal dimensions, continuous efforts have been devoted to examining residential segregation, but mainly from the perspective of race or ethnicity (Johnston et al., 2005; Musterd and Van Kempen, 2009). In these studies, social groups were modeled as binary or categorical variables (for example, white/black people). Thus, many commonly used segregation indices, such as index of dissimilarity (Duncan and Duncan, 1955), isolation index (White, 1986), and Theil's entropy index (Theil, 1972), can be directly applied to measure segregation patterns within a population or region. These indices, although frequently used, are not without limitations. First, many existing indices (for example, index of dissimilarity and isolation index) only provide the overall segregation within a given region or population, thus failing to distinguish the differences among individuals or places. Another critique is that traditional methods usually focus on measuring groups that can easily be defined as nominal categorical variables (Reardon, 2009). This makes them inapplicable when social groups need to be modeled as ordered categories or as a continuous variable. For instance, most of the existing methods cannot deal with income segregation, which is an important dimension of segregation in many contemporary cities.

### **Beyond Residential Segregation: Toward a Dynamic View**

Through the years, researchers have gained numerous insights into segregation at places of residence. With the help of longitudinal census data, residential segregations in many cities and their evolutions have been well documented. It is found that residential income segregation—in twenty-seven of the thirty largest metropolitan areas in the United States—has increased during the past three decades (Fry and Taylor, 2012). Besides separating the rich from the poor, these cities are also producing an increased isolation between racial groups (Fry and Taylor, 2012). Similarly, there has been an increase in socioeconomic segregations in European capital cities between 2001 and 2011 (Musterd et al., 2017). Such an increase, as



concluded by the authors, is linked with structural factors such as social inequalities, globalization and economic reconstructing, welfare regimes, and housing systems.

Much has been investigated about residential segregation. But what about other places? How are people brought together in cities when they engage in different types of activities? Some of the efforts dedicated to these questions include the work from Ellis et al. (2004), Hellerstein and Neumark (2008), and Åslund and Skans (2010), in which they investigated racial or ethnic segregation at workplaces. Two key messages are delivered in these studies. First, besides a spatial separation of social groups at people's residence, some US and European cities also exhibit substantial workplace segregation (Hellerstein and Neumark, 2008; Åslund and Skans, 2010), indicating other factors that would influence the social mixings in cities (for example, labor market, immigration policy, and education). On the other hand, it is shown that workplaces could reduce segregation by bridging people with different social backgrounds (Ellis et al., 2004). This suggests the importance of pursuing a dynamic view of social segregation in cities.

## **In twenty-seven of the thirty largest metropolitan areas in the United States residential income segregation has increased during the past three decades. Besides separating the rich from the poor, these cities are also producing an increased isolation between racial groups**

As we acknowledge the contributions from these researchers, one thing worth mentioning is that their results are still delivered through snapshots at certain locations (for example, workplaces). The diurnal patterns of social segregation in cities remain largely untapped. Since census data only depict a static view of population distribution, they are not suitable for measuring segregation beyond where people live and work. Travel surveys can capture movements and demographic information of the same population. Such data could enrich the temporal aspects of segregation in cities. In one recent study, Le Roux et al. (2017) used a travel survey data to quantify segregation hour by hour from respondents' educational and socio-professional indicators. The authors found that "segregation within the Paris region decreases during the day and that the most segregated group (the upper-class group) during the night remains the most segregated during the day" (p. 134, Le Roux et al., 2017). Such a new perspective of "segregation around the clock" reveals the dynamic impacts of urban spaces on social mixings, which could also inspire new practices in urban design.

### **Urban Segregation in the ICT and Big Data Era: Toward A Hybrid View**

On April 3, 1973, Martin Cooper at Motorola used a handheld device to make the first ever mobile telephone call in history. Since then, information and communication technologies (ICTs) have developed rapidly, unifying telephone and computer networks into an interconnected system. As a collective benefit for modern society, the social channels for human interactions have been greatly enriched. Face-to-face communication is no longer the only way to maintain interpersonal relationships in the contemporary world. Mobile phones, e-mails, and various social networking sites have become the new darlings for



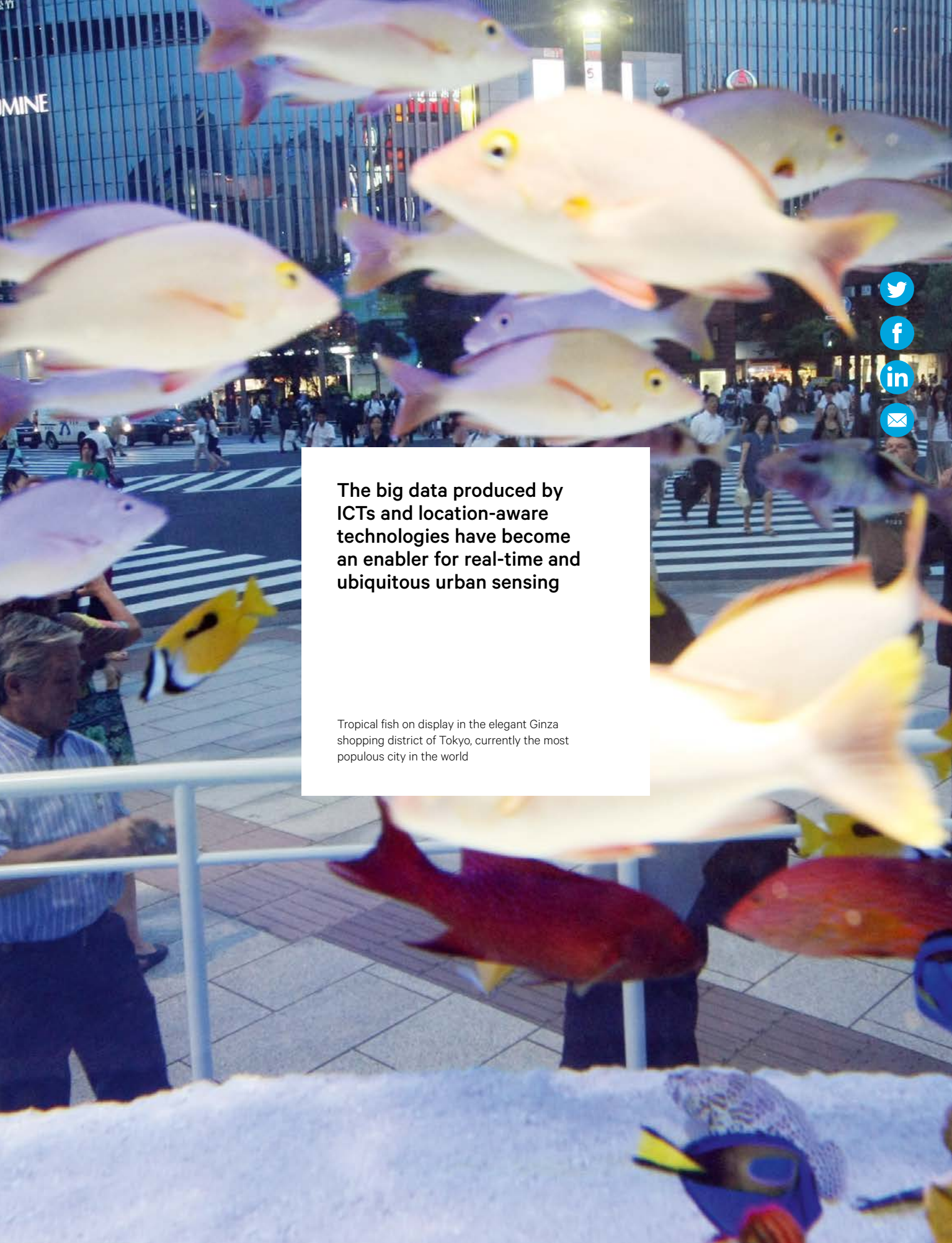
human social interactions. Such technological advancements have profound yet unclear implications on societies' social structures. On the one hand, some people believe that human social interactions are no longer bounded by where they live, work, and entertain. In other words, geographic distance or proximity are expected not to influence how people connect with others. On the other hand, evidence has shown that the structures of online social networks are potentially shaped by the physical world. One good example is the “distance decay” effect discovered in many online social networks, in which the probability of friendship is inversely proportional to the geographic distance (Krings et al., 2009; Xu et al., 2017). That means that human interactions in physical and social space are related to each other. Unifying the two spaces, therefore, would result in a more comprehensive picture of social mixings in cities.

The past couple of years have witnessed some new efforts in studying urban segregations. For example, Silm and Ahas (2014) conducted a temporal analysis of ethnic segregation in Tallinn, Estonia, based on human movements extracted from mobile-phone data. The study demonstrates a cost-effective approach that can capture the diurnal patterns and long-term evolution of urban segregation. Mobile-phone data, which are usually collected by cellular companies for billing purposes, were put under the spot light in this study. Due to the abilities to capture movements of large populations, such data have provided new opportunities for urban dynamics research. Besides human mobility, mobile-phone data can also reveal social network structures. In a more recent study, Leo et al. (2016) examined the socioeconomic imbalances in human interactions using a coupled dataset that records mobile-phone communications and bank transactions of users in a Latin American country. The authors concluded that people tend to be better connected to others of their own socioeconomic classes. These studies demonstrate the potentials of big data in revealing urban segregation—in both physical and social space.

In the past couple of years, Internet and mobile technologies have proliferated around the globe. By 2016, the LTE networks have reached almost four billion people on the earth, covering 53% of the global population (Sanou, 2016). In many developed and developing countries, the mobile penetration rates have reached over one hundred percent. Such numbers are expected to be even higher in urban areas. We are moving toward an age where most of the people use mobile phones to connect to the outside world. The big data produced by ICTs and location-aware technologies have become an enabler for real-time and ubiquitous urban sensing. Mobile-phone data, for example, can be used to analyze city-scale population dynamics. If coupled with sociodemographic information, such data can provide a timely view of social mixings in cities. What is more important is that mobile-phone data can explicitly capture human communications at the individual level. Such information can provide direct evidence on the degree of “homophily” in social networks, which is a salient dimension of segregation in cities.

Note that cities are paying attention to the social merit of these data. Recently, the Türk Telekom initiated a big data challenge named D4R (<http://d4r.turktelekom.com.tr>). In this initiative, selected research groups will be granted access to an anonymized mobile-phone dataset, so that innovative solutions can be proposed to improve the living conditions of Syrian refugees in Turkey. The dataset, which captures the sightings of refugees and non-refugees when they use mobile phones, allows researchers to analyze their movement and interaction patterns. Such data makes it possible to examine how well the refugees integrate with the others in the country, which would ultimately affect their education, employment, safety, and health conditions. As we celebrate the emergence of these voluminous datasets, one question worth asking is how we can come up with novel metrics—by leveraging rich urban

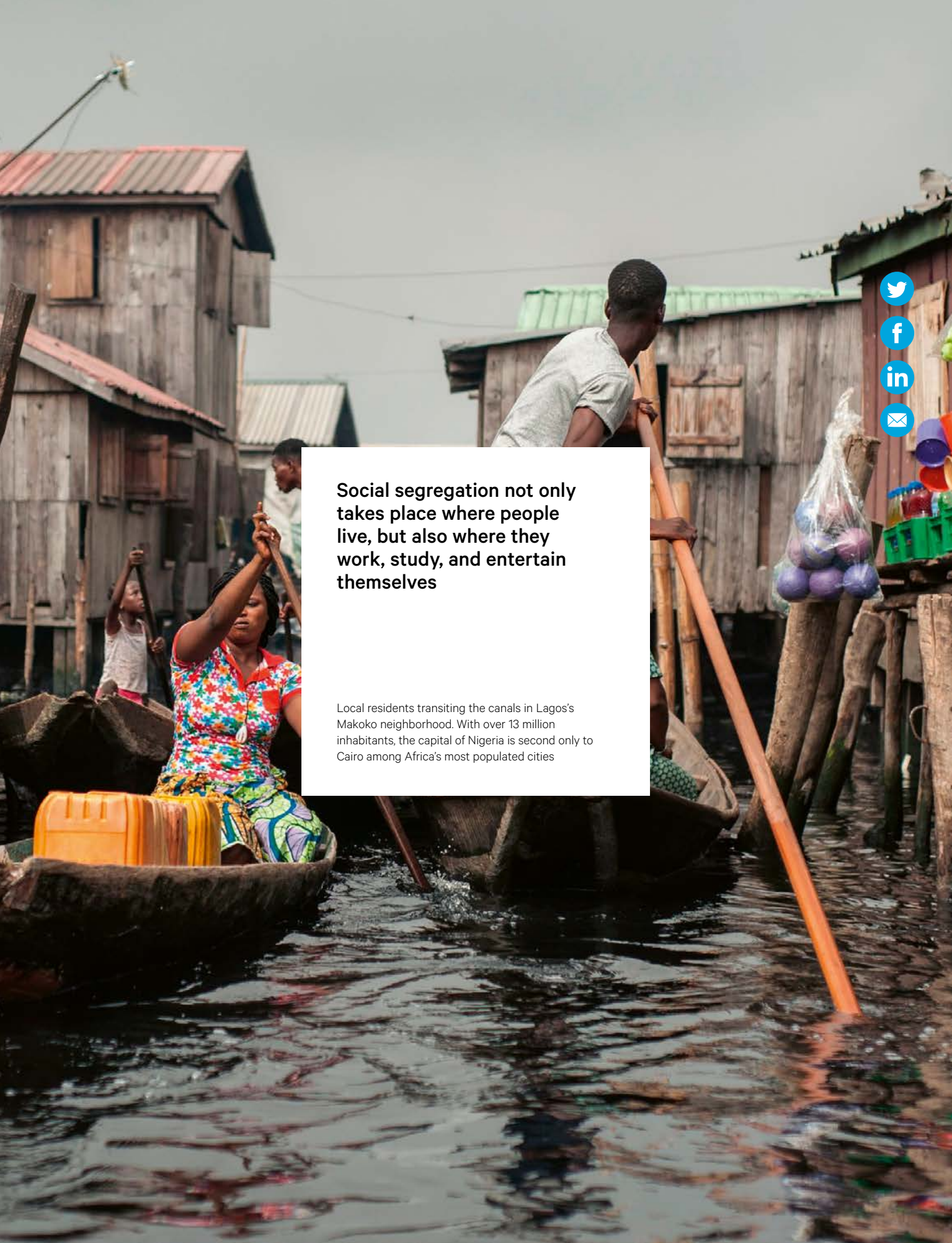




**The big data produced by ICTs and location-aware technologies have become an enabler for real-time and ubiquitous urban sensing**

Tropical fish on display in the elegant Ginza shopping district of Tokyo, currently the most populous city in the world





**Social segregation not only takes place where people live, but also where they work, study, and entertain themselves**

Local residents transiting the canals in Lagos's Makoko neighborhood. With over 13 million inhabitants, the capital of Nigeria is second only to Cairo among Africa's most populated cities

datasets—to reveal the presence or absence of segregation in cities, and how such insights can inform new urban practices?



### Prospect on a New Metrics for Urban Segregation

By reviewing existing segregation studies, we feel that it is time to think about a new metrics that can deliver a dynamic and multidimensional view of segregation in cities. But what features should this metric have? What things need to be measured? We start the reflection by referring to some of the research gaps in this field. First, previous studies have put a considerable focus on measuring residential segregation. Census data, which are often used in these studies, provide a static view of population footprint. They are expensive to collect and their update cycles are very long (for example, five or ten years). This makes it difficult to capture a timely view of social mixings in cities. Although travel surveys have mitigated this issue by providing fine-grained human activity information, collecting such data is still costly and time-consuming. We are in need of cost-effective ways to enrich the spatial and temporal aspects of segregation in cities. Second, current measures are mainly designed for measuring spatial segregation. As part of the human interactions become “virtual,” we need new ways to quantify segregation in social space as well as its relationship with the physical counterpart. Third, an improved understanding is needed on how well individual citizens are integrated within cities. Although individual movements and communication activities can now be captured through various sensing technologies, such data have not been used to quantify segregation at the individual level. Thus, we think a good metric for urban segregation should be able to fulfill all or at least some of the following objectives:

- to quantify urban segregation in both physical and social space;
- to depict temporal variation of segregation in cities;
- to distinguish segregation patterns at different urban locations;
- to measure different types of segregations (for example, racial segregation, ethnic segregation, and income segregation);
- to deliver individual-level segregation measures;
- to support intra- and inter-city comparisons.

If equipped with such capabilities, the metrics can then be used to answer profound questions about urban segregation, such as:

- which places in a city, and at what time, are mainly occupied by similar socioeconomic classes? Which places are used by a great diversity of social groups?
- what types of places contribute most to the social integration of a city?
- to what extent is an individual exposed to similar others in physical space or social space? Is there a correlation between the two? (that is, are people who are more isolated in physical space also more segregated in social space?);
- which cities are experiencing more severe social segregations than others?
- how are urban segregations related to the characteristics of cities (for example, economic development, demographic composition, migration patterns, and urban form)?

Figure 1 presents our perspective on a new metrics for urban segregation. As the underpinning of the metrics, multi-source urban data provide new opportunities for coupling movements,



social interactions, and demographic characteristics of large populations. Different data types—with their own pros and cons—can be combined in an organic way to facilitate urban segregation studies. Mobile-phone and social-media data, for example, can capture human movements and social interactions simultaneously. Such data could support mobility and social network analysis at population scales. However, due to privacy concerns, these data usually fall short of collecting individual sociodemographic characteristics. Such information, which is important to segregation studies, can be inferred through other innovative approaches. It is found that behavioral indicators extracted from mobile-phone usage can accurately predict an individual's socioeconomic status (Blumenstock et al., 2015). Other urban datasets, such as census and residential property price, also contain valuable information about urban residents. As a person's frequented activity locations (for example, home) can be identified from human activity data (for example, mobile-phone data), such information can be further associated with one's home census tract or local housing price to establish his/her socioeconomic profile. Utilizing such prediction and data fusion techniques allows us to gather valuable input for segregation analysis.

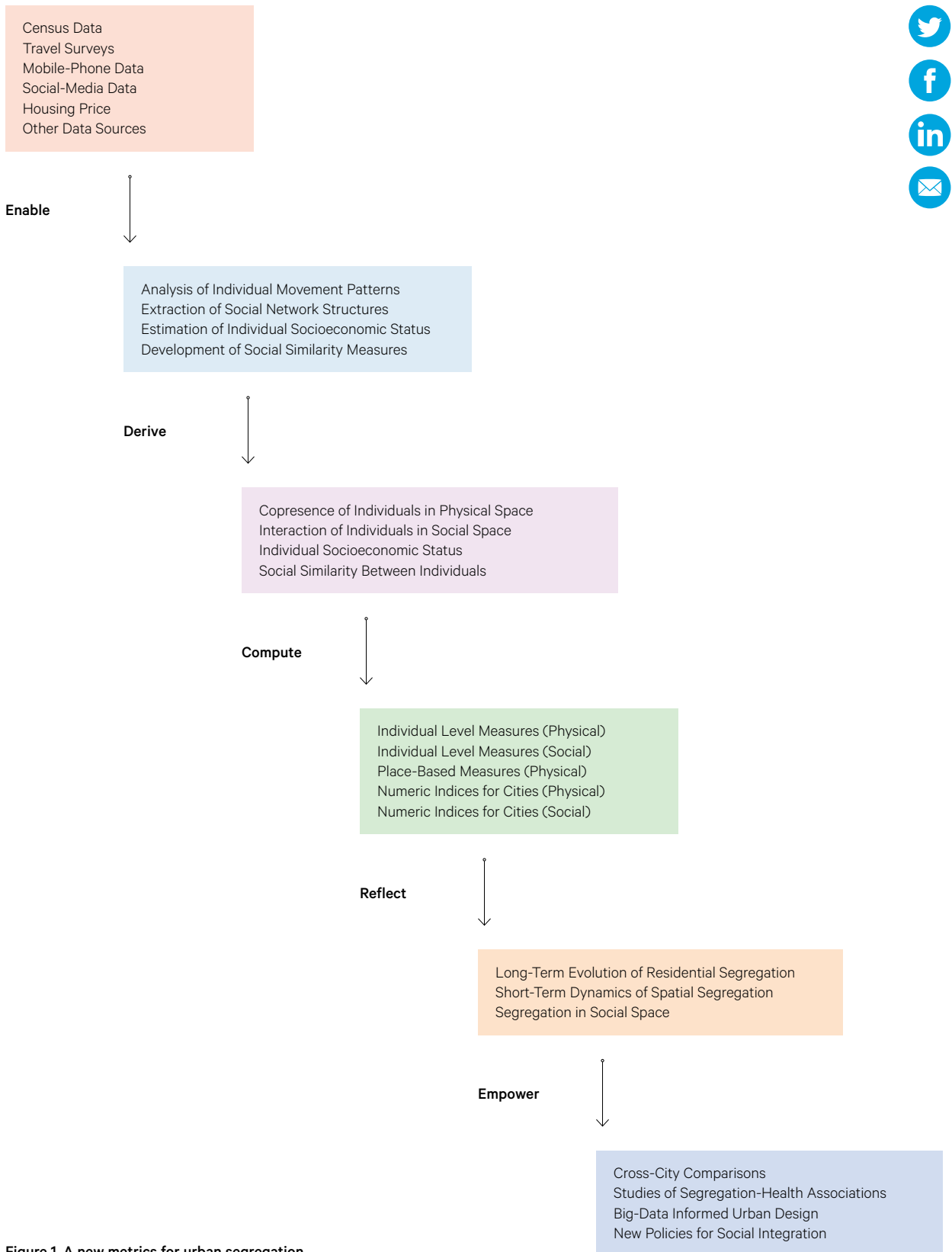
**It is found that behavioral indicators extracted from mobile-phone usage can accurately predict an individual's socioeconomic status. Other urban datasets, such as census and residential property price, also contain valuable information about urban residents**

Once such information is available, the next step is to build segregation indicators. But how to formulate measures in a quantitative and reasonable way? Human movement patterns extracted from the previous step (light-blue tier in fig. 1) can be used to quantify the *co-presence* of individuals in physical space. The measure quantifies how likely two given individuals tend to appear at the same location at approximately the same time. Combining this information with a social similarity metric could effectively describe how individuals are segregated in physical space. For example, an individual who spends most of his time at locations that are mainly occupied by similar others is considered as physically segregated. A similar measure can be designed to quantify segregation in the social space. People who are segregated in social space are those who are dominated by connections (for example, frequent phone communications) within their own socioeconomic classes.

Other than these two individual-level measures (green tier in fig. 1), we can also establish place-based measures. For example, a place which is always used by people with similar social characteristics is quite different from one that is used by a variety of social groups. Such a measure can be realized by first establishing a vector at a given location—with each element representing the proportion of observed people belonging to a certain social class—followed by the calculation of the entropy of this vector. The measures can also be designed as time dependent so that the dynamic characteristics of urban spaces can be quantified.

Upon measures that distinguish individual citizens or places, we also need numeric indices to describe the overall segregation of cities so that they can be compared. On the one hand, existing indices—such as index of dissimilarity and Theil's entropy index—can be applied directly to quantify segregation in physical space. On the other hand, very few efforts, if any, have been devoted to quantifying the overall segregation in social space (for example, in a mobile social network). There is a need to design new indices to facilitate this purpose. One





possible approach would be to measure the deviation from the observed segregation patterns from a “null model,” which assumes that interactions between people are established randomly in social space.

With these new measures, we can monitor not only the long-term evolution, but also the short-term dynamics of urban segregation (orange tier in fig. 1). Our observations will be transformed from physical space to a hybrid physical-social space. The ways we quantify segregation will be enriched to include not only aggregate measures, but also individual-level measures. Such a metrics could generate new momentum for academic research and, more importantly, produce numerous social benefits (dark-blue tier in fig. 1).



### Shaping Cities for Social Integration

Social segregation is deeply rooted in many cities worldwide, and it can become more pronounced in the future as urbanization accelerates. Overcoming this issue, however, has never been easy. Factors that breed urban segregations are complicated, usually involving historical, political, and economic forces that are closely intertwined. The emerging urban datasets and the proposed metrics, as we believe, could yield a more comprehensive view of socioeconomic mixings in cities. This can not only inspire new practices for improving urban integration, but also help evaluate the effectiveness of existing measures.

Although the ways of quantifying segregations are becoming more diverse, we should always bear in mind the persistent impact of residential segregation on societal well-being. The “orthodox” urban planning, as criticized by American activist Jane Jacobs, results in zoning restrictions and gentrification that tear urban societies apart. The metrics proposed in this article could empower decision-makers to better understand the behavioral dynamics of people who live in socially isolated areas, for example, to assess their segregation levels during daytime as well as in social space. This will move us beneath the surface of residential segregation, from which new policies and intervention strategies can be developed.

Creating diverse public spaces that bridge different socioeconomic classes is something cities should pursue. Building walkable cities with small street blocks, as advocated by Jane Jacobs, could possibly facilitate social interactions while reinforcing informal surveillance by the “eyes on the street.” With this new metrics, we can analyze different design principles—by correlating place-based segregation measures with built environment indicators (for example, block size, road network density, and land use diversity)—to evaluate the impact of public spaces on social integrations.

**Creating diverse public spaces that bridge different socioeconomic classes is something cities should pursue. Building walkable cities with small street blocks could possibly facilitate social interactions while reinforcing informal surveillance by the “eyes on the street”**

On the mobility side, cities could consider building transportation links that cover areas that are less easily accessible. The cable-car system implemented in Medellín, Columbia, is a successful example that improves the connections between the city’s informal settlements and

the others. The improvements in transit services could strengthen the interactions among urban locations, which enhance the location choices of people, especially the disadvantaged groups. Of course, such improvements should come along with relevant policies, for example, providing affordable housing to the less affluent population and creating mixed-income neighborhoods.

As Internet and mobile technologies proliferate, facilitating online social integration has become more important than ever before. Regulations should be carried out to ensure that online information (for example, job opportunities) is equally accessible to different social groups, and, meanwhile, to create virtual communities that bridge people of different kinds.

Our world is experiencing rapid urbanization. Behind the fast development of contemporary cities, our urban environments are confronted with tremendous social and economic inequality. Integration is at the core of cities since their emergence over 10,000 years ago. Only if we are able to measure to what extent this primordial function still performs and—when needed—implement correct measures can we move toward a safe urban future.



## Acknowledgments

We acknowledge support from the Research Grant Council of Hong Kong (project no. 25610118) and the Hong Kong Polytechnic University Start-Up Grant (project no. 1-BE0J).

## Select Bibliography

- Åslund, O., and Skans, O. N. 2010. "Will I see you at work? Ethnic workplace segregation in Sweden, 1985–2002." *ILR Review* 63:3, 471–493.
- Blumenstock, J., Cadamuro, G., and On, R. 2015. "Predicting poverty and wealth from mobile-phone metadata." *Science* 350:6264, 1073–1076.
- Cortese, C. F., Falk, R. F., and Cohen, J. K. 1976. "Further considerations on the methodological analysis of segregation indices." *American Sociological Review* 630–637.
- Duncan, O. D., and Duncan, B. 1955. "A methodological analysis of segregation indexes." *American Sociological Review* 20:2, 210–217.
- Ellis, M., Wright, R., and Parks, V. 2004. "Work together, live apart? Geographies of racial and ethnic segregation at home and at work." *Annals of the Association of American Geographers* 94:3, 620–637.
- Fry, R., and Taylor, P. 2012. "The rise of residential segregation by income." *Washington, DC: Pew Research Center* (2012), 26.
- Hellerstein, J. K., and Neumark, D. 2008. "Workplace segregation in the United States: Race, ethnicity, and skill." *The Review of Economics and Statistics* 90:3, 459–477.
- Johnston, R., Poulsen, M., and Forrest, J. 2005. "On the measurement and meaning of residential segregation: A response to Simpson." *Urban Studies* 42:7, 1221–1227.
- Krings, G., Calabrese, F., Ratti, C., and Blondel, V. D. 2009. "Urban gravity: A model for inter-city telecommunication flows." *Journal of Statistical Mechanics: Theory and Experiment* 2009:07, L07003.
- Le Roux, G., Vallée, J., and Commenges, H. 2017. "Social segregation around the clock in the Paris region (France)." *Journal of Transport Geography* 59: 134–145.
- Leo, Y., Fleury, E., Alvarez-Hamelin, J. I., Sarraute, C., and Karsai, M. 2016. "Socioeconomic correlations and stratification in social-communication networks." *Journal of The Royal Society Interface* 13:125, 20160598.
- Massey, D. S., and Denton, N. A. 1987. "Trends in the residential segregation of Blacks, Hispanics, and Asians: 1970–1980." *American Sociological Review* 802–825.
- Massey, D. S., and Denton, N. A. 1988. "The dimensions of residential segregation." *Social Forces* 67:2, 281–315.
- Musterd, S., Marcińczak, S., Van Ham, M., and Tammaru, T. 2017. "Socioeconomic segregation in European capital cities. Increasing separation between poor and rich." *Urban Geography* 38:7, 1062–1083.
- Musterd, S., and Van Kempen, R. 2009. "Segregation and housing of minority ethnic groups in Western European cities." *Tijdschrift voor economische en sociale geografie* 100:4, 559–566.
- Reardon, S. F. 2009. "Measures of ordinal segregation." In *Occupational and Residential Segregation*, Yves Flückiger, Sean F. Reardon, and Jacques Silber (eds.). Bingley: Emerald Publishing, 129–155.
- Reardon, S. F., and O'Sullivan, D. 2004. "Measures of spatial segregation." *Sociological Methodology* 34:1, 121–162.
- Sanou, B. 2016. *ICT Facts and Figures 2016*. International Telecommunication Union.
- Silm, S., and Ahas, R. 2014. "The temporal variation of ethnic segregation in a city: Evidence from a mobile-phone use dataset." *Social Science Research* 47: 30–43.
- Taeuber, K. E., and Taeuber, A. F. 1976. "A practitioner's perspective on the index of dissimilarity." *American Sociological Review* 41:5, 884–889.
- Theil, H. 1972. *Statistical Decomposition Analysis: With Applications in the Social and Administrative Sciences* (No. 04; HA33, T4).
- United Nations. 2014. *World Urbanization Prospects: The 2014 Revision, Highlights*. Department of Economic and Social Affairs, Population Division, United Nations.
- White, M. J. 1986. "Segregation and diversity measures in population distribution." *Population Index* 198–221.
- Williams, D. R., and Collins, C. 2001. "Racial residential segregation: A fundamental cause of racial disparities in health." *Public Health Reports* 116:5, 404.
- Xu, Y., Belyi, A., Bojic, I., and Ratti, C. 2017. "How friends share urban space: An exploratory spatiotemporal analysis using mobile-phone data." *Transactions in GIS* 21:3, 468–487.





**Amos N. Guiora**  
SJ Quinney College of Law,  
University of Utah

Amos N. Guiora is Professor of Law at the S. J. Quinney College of Law, University of Utah. He teaches Criminal Procedure, Criminal Law, Global Perspectives on Counterterrorism and Religion and Terrorism, incorporating innovative scenario-based instruction to address national and international security issues and dilemmas. Professor Guiora is a Research Associate at the University of Oxford, Oxford Institute for Ethics, Law and Armed Conflict; a Research Fellow at the International Institute for Counter-Terrorism, The Interdisciplinary Center, Herzliya, Israel; a Corresponding Member, The Netherlands School of Human Rights Research, the Utrecht University School of Law; and Policy Advisor, Alliance for a Better Utah. Professor Guiora has published extensively both in the US and Europe. He is the author of several books and book chapters, most recently: *Cybersecurity: Geopolitics, Law, and Policy*; *The Crime of Complicity: The Bystander in the Holocaust* (translated into Chinese and Dutch), second edition forthcoming; and *Earl Warren, Ernesto Miranda and Terrorism*. Professor Guiora has been deeply involved over a number of years in Track Two negotiation efforts regarding the Israeli-Palestinian conflict predicated on a preference and prioritization analytical tool.

Recommended book: *Cybersecurity: The Essential Body of Knowledge*, Dan Shoemaker and W. Arthur Conklin, Cengage, 2011.

The degree of gravity with which we view the cyber threat is a matter of perspective, experience, interaction, and awareness. Whether we view it as a criminal act or terrorism or a hybrid is an important point for discussion. Resolving that question helps frame how we perceive those responsible for cyberattacks and facilitates resolution regarding punishment for their actions. That question has relevance for law enforcement, prosecutors, defense attorneys, and the judiciary. Our focus, as the title suggests, is on *cybersecurity cooperation* in an effort to minimize the impact of an attack. Cooperation, as we shall come to see, raises concerns in different quarters. Some of that concern is legitimate. It needs to be addressed and compellingly responded to. A cooperation model is intended to offer institutionalized mechanisms for preventing—or at least minimizing—the harm posed by the mere threat of an attack, much less an attack itself.



It is commonly accepted today that cyber threat is a reality. The degree of gravity with which we view it is a matter of perspective, experience, interaction, and awareness. Whether we view it as a criminal act or terrorism or a hybrid is an important point for discussion.

Resolving that question helps frame how we perceive those responsible for cyberattacks and facilitates resolution regarding punishment for their actions. That question has relevance for law enforcement, prosecutors, defense attorneys, and the judiciary. There are also moral considerations relevant to this discussion in the context of responsibility and accountability.

An additional reality, historically, is a lack of cooperation among distinct—yet interconnected—communities that have much to gain by banding together in the face of a grave threat. A retrospective analysis of how potential targets of cyberattacks, and those with a vested interest—and capability—to minimize the threat, fail to join forces paints a disconcerting picture. That picture, in a nutshell, reflects an unwillingness—for a variety of reasons—to recognize the benefits of a unified response to cyberattacks.

This reluctance or hesitation is not new to 2018; my research while engaged in cybersecurity writing projects made that very clear. In other words, the lack of cooperation noticed in 2018 is not a recent development, quite the opposite: a review of the relevant data, literature, and anecdotal evidence over the years convincingly highlights an—for lack of a better term—institutionalized lack of cooperation. In considering different reasons for the success enjoyed by cyber actors, their nefariousness is enhanced by a consistent failure of potential targets to recognize the benefits of adopting a cooperation model. Given the persistent, committed, and unrelenting efforts of those committed to engaging in cyberterrorism/cyberattacks, the failure to consistently commit to cooperation models is noteworthy, and troubling. Noteworthy because it reflects a consistent pattern; troubling because it facilitates continued harm to society, broadly defined.

## **Cyber threat has relevance for law enforcement, prosecutors, defense attorneys, and the judiciary. There are also moral considerations relevant to this discussion in the context of responsibility and accountability**

That historical pattern explains, in part, the success of cyberattacks. It reflects a failure to fully develop defensive measures that would minimize, if not mitigate, the harm perpetuated by the effects of cyberattacks. Recognizing the consequences of a lack of cooperation stands out when reviewing cybersecurity issues over the years; developing, and implementing, cooperation models reflects a positive—and much-needed—measure to more effectively respond to a clear threat. Those twin realities are at the core of the discussion in the pages ahead.

Framing an issue has importance in determining resources allocated to countering, mitigating, and minimizing threats. Similarly, resolving the terrorism–criminal law–hybrid dilemma is relevant in determining operational measures that can be implemented against the actor, or actors, responsible for a particular attack. However, as important as that discussion is, it is not the one pursued in the pages that follow. Our focus, as the title suggests, is on *cybersecurity cooperation* in an effort to minimize the impact of an attack. Cooperation,



as we shall come to see, raises concerns in different quarters. Some of that concern is legitimate. It needs to be addressed and compellingly responded to.

To undertake a discussion regarding cybersecurity requires defining the term. As is always the case in such matters, reasonable minds can reasonably disagree. That is a given and understandable.

I define cyberattacks as a “deliberate, nefarious use of technology intended to harm individuals, communities, institutions, and governments.” While this proposed definition is subject to disagreement, it is assumed that “harm”—regardless of whether an attack was successful—is a largely agreed-upon notion.

For cyber to be effective, there need not be an attack, nor must it be successful. Given the omnipresence of cyber—the mere fear of such an attack—we are, individually and collectively, spending significant resources, time, and energy on minimizing the potential impact of an attack. Much like counterterrorism, significant efforts are allocated to both preventing an attack and, were one to occur, to its impact and consequences. The point of inquiry is how do we most effectively protect ourselves; are we doing so presently; and are there more effective, creative mechanisms for doing so. It is safe to assume that those same questions are asked, in the reverse, by those dedicated to deliberately harming civil society, broadly and narrowly defined.

## **Instituting cooperation models intended to offer institutionalized mechanisms will help to prevent, or at least minimize, the harm posed by the mere threat of an attack, much less an attack itself**

That is the case regardless of the paradigm assigned to cyber. Nevertheless, the importance—as referenced above—of determining the definition and category alike of cyber cannot be minimized. However, as important as that is, and its importance must not be gainsaid, our focus is on a different aspect of cyber.

To that, we turn our attention.

In a nutshell, it is cooperation among distinct potential targets and among different law-enforcement agencies and between targets and law enforcement. Concisely stated, it proposes that the impact of a cyberattack can be minimized by instituting cooperation models. It is intended to offer institutionalized mechanisms to prevent—or at least minimize—the harm posed by the mere threat of an attack, much less an attack itself. Given the consequences of an attack, and the resources required to minimize its impact, the proposed cooperation model is intended to minimize costs, direct and indirect alike, of cyber. The model is predicated on an assumption: preventing an attack, or at the very minimum making a concerted-determined effort, is preferable to absorbing the costs of a successful attack.

As I have come to learn, there are corporate executives who subscribe to a distinctly different argument: it is more cost-effective to absorb a hit rather than invest in sophisticated protection models. I have met with executives whose corporations were the targets of successful attacks who subscribe to this theory.<sup>1</sup> For me, this rationale reflects short-term thinking focused on narrow economic considerations, suggesting a lack of understanding regarding benefits accruing from proactive protection and cooperation models. These people also articulated concerns regarding public image, customer anxiety, and an upper hand that competitors might gain. From a strictly business perspective, narrowly defined and applied, such an approach



is, perhaps, understandable. The argument made was rational, neither impulsive nor spur of the moment; from my interactions, it was also clear that my interlocutors considered this perspective reflective of best-business practices.

This is something that is not to be casually dismissed with a wave of the hand. Nevertheless, their argument reflects narrow tactical considerations, focused internally, devoid of examining the larger picture and possible benefits that may accrue. Whereas information sharing, in an effort to minimize harm, was presented to these executives as a positive, their (inward looking) perspective led them to perceive information sharing as a negative, with significant impact on their product and positioning.

It is important to add that law-enforcement officials perceived the cooperation model as a positive, enhancing their efforts, albeit with one caveat: clearly articulated hesitation by federal law-enforcement officials to cooperate information share with their state and local counterparts. Whether this is institutionalized hesitation or limited to specific individuals is unclear; exploring this question did not lead to clarity one way or the other.

In this article I hope to convince readers of the shortsightedness of such an approach and to highlight the benefits of the proposed cooperation model.

## For cyber to be effective, there need not be an attack, nor must it be successful

The question, in a nutshell, is: what is the most effective means for enhancing cybersecurity?

To convince the reader of the viability of such a proposal, we must examine the costs of cyber, threat and attack alike. That is, there must be a benefit accrued to acceptance of a cooperation model. That benefit must outweigh any attendant costs, otherwise the proposal is a nonstarter from the very outset. Similarly, concerns must be allayed that cooperation is not an “own goal,” whereby positive intentions have harmful, unintended consequences that mitigate the value of cooperation. The working premise, as discussed below, is that cooperation is a positive, should be viewed accordingly, and that models need to be developed to facilitate its successful implementation.

This will not be easy as was made clear to me both while writing my book, *Cybersecurity: Geopolitics, Law, and Policy*,<sup>2</sup> and then reinforced while conducting a tabletop exercise.<sup>3</sup> Notwithstanding the arguments posed—some understandable, others best characterized as “head scratchers”—I believe that the proposed cooperation model deserves careful attention in the face of relentless cyberattacks. As the cliché goes: “The best defense is a good offense.”

The discussion below will be divided into the following five sections: “What does Cooperation Mean”; “Cooperation Among Whom”; “Cooperation to What Extent”; “The Good, the Bad and the Ugly of Cooperation”; and “Final Word.”

### What does Cooperation Mean

Cooperation has many definitions and meanings. Context, times, culture, and circumstances play an important role in both defining the term and its actual implementation. Cooperation can be permanent, transitory, or situational; it can be predicated on formal arrangements and informal understandings alike. Cooperation, whether bilateral or multilateral, is at its best when the parties believe the relationship is mutually beneficial.





That does not suggest that such an agreement—formal or informal—is cost-free. It is not. But that is the reality of any agreement for the implication is that certain “freedoms” are voluntarily relinquished for a perceived benefit. One has only to recall the writings of Hobbes, Locke, or Rousseau to recognize the costs—and benefits—of creating and joining a community.<sup>4</sup>

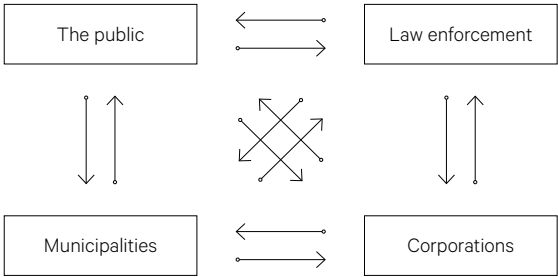
That is the essence of cooperation. In joining a community, and cooperating with its members, we seek safety, security, and protection. That is generally perceived to be a positive; in return, we give up individualism, a certain degree of free will, and some independence. For some, that is an unacceptable cost, not outweighed by benefits that may be procured by the positives that emanate from cooperation. Fair enough.

However, as history suggests, in the main the benefits of cooperation outweigh perceived, or actual, costs. Exceptions exist. Of that there is no doubt. However, the notion of “banding together” for a common enterprise or undertaking is generally perceived to be preferable to “going it alone.” To facilitate—if not enhance—effectiveness, an agreement of cooperation must include:

- 1. consensus regarding common goals;
- 2. mechanisms for assessing continued costs–benefits;
- 3. mechanisms for withdrawing from the agreement;
- 4. established parameters of the agreement;
- 5. agreed-upon implementation mechanisms of the agreement.

Needless to say, the underlying predicate is that parties enter into the agreement voluntarily, recognizing a need for such an arrangement, and recognizing that the alternative is not beneficial. As we move forward in this discussion, these principles serve as the basis for our recommendation that cooperation be an integral aspect of minimizing the threats posed by cyberattacks, whether viewed as a crime, terrorism, or a hybrid.

To put flesh on the bones: the cooperation model I recommend would benefit private and public entities alike. This model would, similarly, require—oblige—corporations that have been hacked to institutionalize reporting mechanisms to law enforcement and customers alike. The intersecting vectors between the public (broadly speaking), municipalities, law enforcement, and corporations (public and private alike) are numerous.



This complexity reflects the multiplicity of threats and vulnerable targets alike. To implement this recommendation is, as I have come to learn, counterintuitive for many affected parties, including corporations (not surprisingly) and law enforcement (surprisingly). The notion, as discussed below, of perceiving this as a collective security model combined with legitimate self-defense (both adopted from international law) was met with skepticism.



Corporate resistance reflected concern regarding potential loss of revenues and unintended consequences in the guise of advantages gained by competitors; law-enforcement hesitation was premised on two different rationales: an unwillingness among federal law-enforcement officials to cooperate with local officials and the lack of resources available to address cyberattacks. The first argument is a nonstarter; the second requires recognition that cybercrimes demand sufficient allocation of resources to position law enforcement as a more effective partner in what must be a collective effort to minimize the threat posed by nefarious actors (see below for a discussion of the Cybersecurity Information Sharing Act, 2015).

How to achieve the corporation model? There are a number of alternatives and options, some reflecting voluntary measures, others legislatively imposed. In the ideal, the former would “rule the day”; however, as was made clear in numerous conversations, that is, unfortunately, unlikely. Which gives rise to a proposal to *legislatively mandate cooperation models*, imposing a duty on potentially impacted targets of cyberattacks (the list of potential targets is outlined under “Cooperation Among Whom,” below).

A word of caution: suggesting congressional intervention, beyond existing requirements, is a source of great discomfort, particularly for the private sector and ideological concerns, and particularly for libertarians who strongly reject such measures. Both are legitimate, and potent, stumbling blocks in seeking to advance such legislation. This potency is magnified by the reality of politics, particularly in the divisive environment that presently defines the US.

## Corporate resistance reflects concern regarding potential loss of revenues and unintended consequences in the guise of advantages gained by competitors

Reporting requirements, as illustrated by the Sarbanes–Oxley Act of 2002,<sup>5</sup> are the bane of many corporate leaders. One corporate executive described the legislation as imposing unwieldy burdens, costs, and obligations on corporations. According to this executive, his team was comprised more of compliance officers than production officers.

Without doubt, that imposes unwanted—perhaps unwarranted—costs on corporations. Nevertheless, the risks and threats posed by cyberattacks justify exploring mechanisms obligating cooperation. That is, if voluntary cooperation is not possible, then the discussion needs to be expanded to the possibility that cooperation be mandated.

The 2015 Cybersecurity Information Sharing Act (S. 754), introduced by US Senator Richard Burr (R-NC) and signed by President Obama on December 18, 2015, was an important measure for it “authorizes cybersecurity information sharing between and among the private sector; state, local, tribal, and territorial governments; and the Federal Government.”<sup>6</sup> The legislation sought to accomplish four important goals; namely it:

- requires the federal government to release periodic best practices. Entities will then be able to use the best practices to further defend their cyber infrastructure;
- identifies the federal government’s permitted uses of cyber threat indicators and defensive measures, while also restricting the information’s disclosure, retention and use;
- authorizes entities to share cyber-threat indicators and defensive measures with each other and with DHS [the Department of Homeland Security], with liability protection;
- protects PII [Personally Identifiable Information] by requiring entities to remove identified PII from any information that is shared with the federal government. It requires that any federal agency that receives cyber information containing PII to protect the PII from unauthorized use

or disclosure. The U.S. Attorney General and Secretary of the Department of Homeland Security will publish guidelines to assist in meeting this requirement.<sup>7</sup>



In the same vein, on July 6, 2016, the European Parliament adopted the Directive on Security of Network and Information Systems (the NIS Directive), which provides legal measures to boost the overall level of cybersecurity in the EU by ensuring:

- Member States' preparedness by requiring them to be appropriately equipped, e.g. via a Computer Security Incident Response Team (CSIRT) and a competent national NIS authority;
- cooperation among all the Member States, by setting up a cooperation group, in order to support and facilitate strategic cooperation and the exchange of information among Member States. They will also need to set a CSIRT Network, in order to promote swift and effective operational cooperation on specific cybersecurity incidents and sharing information about risks;
- a culture of security across sectors which are vital for our economy and society and moreover rely heavily on ICTs, such as energy, transport, water, banking, financial market infrastructures, healthcare and digital infrastructure. Businesses in these sectors that are identified by the Member States as operators of essential services will have to take appropriate security measures and to notify serious incidents to the relevant national authority. Also key digital service providers (search engines, cloud computing services and online marketplaces) will have to comply with the security and notification requirements under the new Directive.<sup>8</sup>

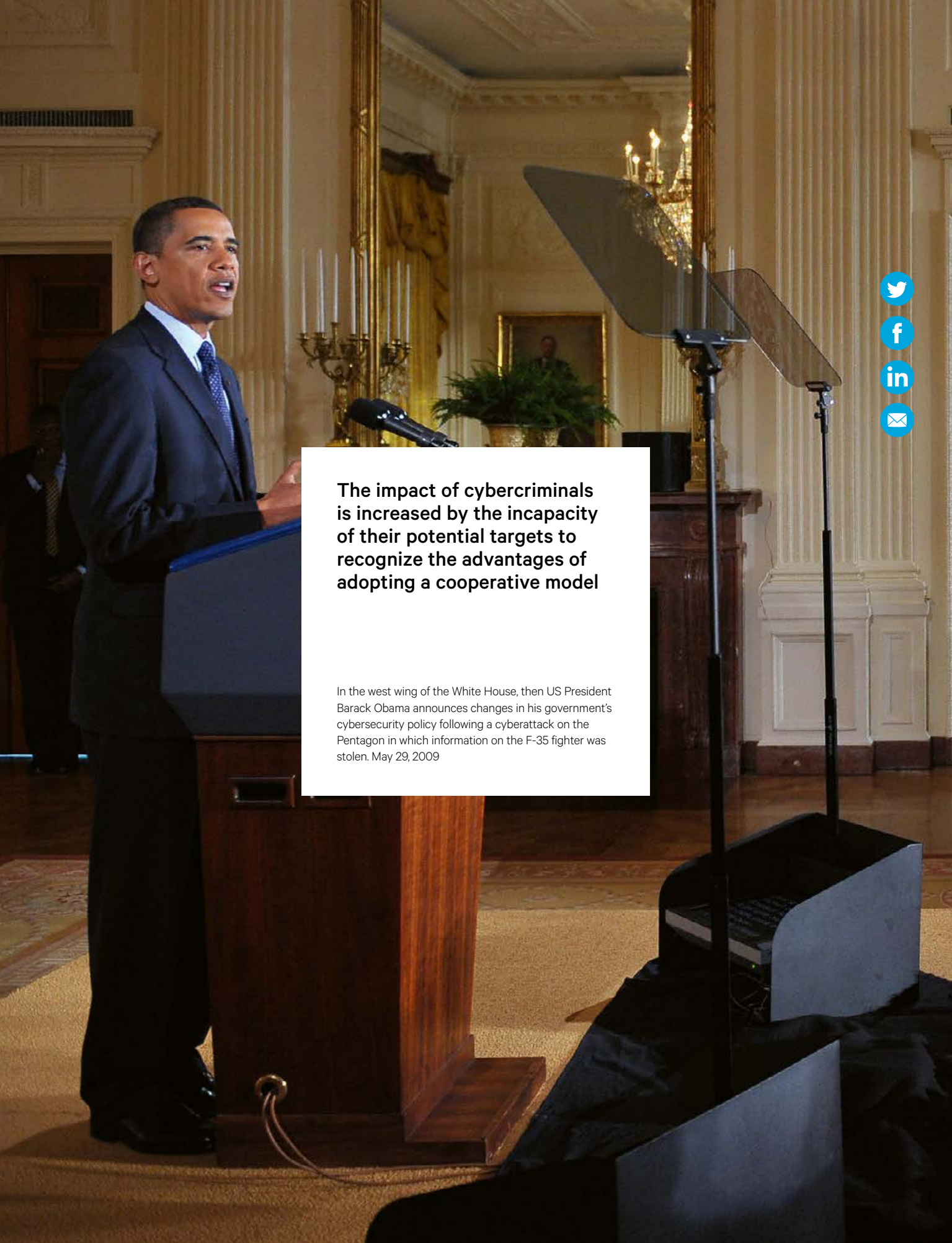
The question is how to build off both these important measures and increase the levels of cooperation among different communities, creating an infrastructure that most effectively minimizes the possible consequences of a cyberattack. This requires integrating sectors, vectors, and interest groups. To do so most effectively requires recognition that existing cooperation measures, as well-intentioned as they maybe, fall short. The “why” was made clear to me; the challenge as we move forward in this discussion is determining whose cooperation must be more effectively institutionalized.<sup>9</sup>

It is to that question that we turn our attention.

### Cooperation Among Whom

From the general to the specific. In the cybersecurity realm, the number of targeted, if not impacted, targets include the following (this is not an all-inclusive list):

- individuals who are the targets of identity theft;<sup>10</sup>
- individuals whose privacy is compromised when, for example, health-care providers are hacked;
- corporations (of all sizes) whose data is compromised;
- the financial industry, including banks, investment houses, stock markets;
- municipalities whose infrastructure (that is, traffic signals, water systems) is compromised;
- hospitals, thereby endangering patient welfare and the safety and protection of health records;
- militaries, when weapon systems are compromised;
- airlines and air-traffic-control systems;
- government agencies and offices.




**The impact of cybercriminals is increased by the incapacity of their potential targets to recognize the advantages of adopting a cooperative model**

In the west wing of the White House, then US President Barack Obama announces changes in his government's cybersecurity policy following a cyberattack on the Pentagon in which information on the F-35 fighter was stolen. May 29, 2009







On July 6, 2016, the European Parliament adopted the NIS Directive, which provides legal measures to boost the overall level of cybersecurity in the EU

Fiberoptic and copper Ethernet cables. The Department of Culture, Communication and Sports is working with the British Ministry of the Interior on a white paper to oblige technology giants to disclose how they manage the offensive and illegal material published by their users







This list, while not inclusive, is sufficiently broad to highlight the range of potential cyber-attack targets. It emphasizes the significant number of people, institutions, corporations, and critical infrastructure at risk from a cyberattack. The notion of “risk” is essential to the cooperation discussion; devoid of potential risk, the impetus for entering into an agreement is significantly minimized, if not entirely eviscerated.<sup>11</sup>

However, the reality of cybersecurity is that it is a world of clear, but unknown, unseen, undiscernible threats until the attack actually occurs. The cyber threat must be phrased as “when will an attack occur” rather than “if an attack will occur.”

The list above reflects the broad range of potential targets. It also highlights the scope of secondary victims, impacted by a particular attack. Furthermore, it accentuates the important question of “to whom a duty is owed.” Cooperation must not be viewed as “nice to have,” but rather as a means necessary to protect intended, and unintended, victims alike of a cyberattack. To state the obvious: the number of individuals impacted by a hack into any of the entities listed above is staggering. Not only is the financial cost extraordinary, but the crippling of a municipality, the impact on a hospital, the danger incurred by air travelers, the consequences faced by those requiring services provided by government agencies are literally overwhelming.

**The reality of cybersecurity is that it is a world of clear, but unknown, unseen, undiscernible threats until the attack actually occurs. The cyber threat must be phrased as “when will an attack occur” rather than “if an attack will occur”**

That in and of itself should make cooperation among the above an obvious measure. I propose we consider this proposition through the lens, by extrapolation, of international law. Doing so, enables us to examine cooperation as both self-defense and collective security. In the context of cooperation among potential victims of a cyberattack, this proposed duality provides an intellectual—and perhaps practical—framework. It highlights two important realities: the necessity of self-defense and recognition that joining forces<sup>12</sup> is an effective means of enhancing protection.

Cooperation, or mutual security, does not imply agreement on all issues nor does it suggest a perfect confluence of interests, values, and goals. It does, however, reflect recognition that certain threats, given their possible consequences, warrant finding a sufficient meeting ground even when the parties may have otherwise competing interests. The two principles—self-defense and collective security—can be viewed as complimentary of each other. On the one hand, individual action is justified; on the other, recognition that certain instances require cooperation in order to facilitate protection.

Article 51 of the United Nations Charter states:

Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security. Measures taken by Members in the exercise of this right of self-defence shall be immediately reported to the Security Council and shall not in any way affect the authority and responsibility of the Security Council under the present Charter to take at any time such action as it deems necessary in order to maintain or restore international peace and security.<sup>13</sup>

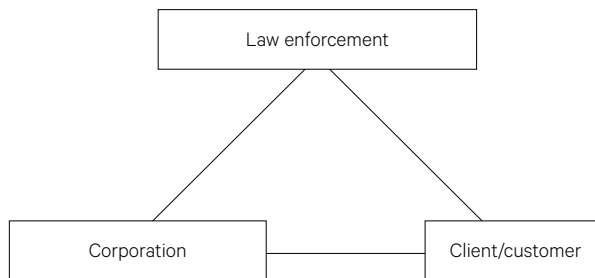
Article 5 of the NATO Treaty states:

The Parties agree that an armed attack against one or more of them in Europe or North America shall be considered an attack against them all and consequently they agree that, if such an armed attack occurs, each of them, in exercise of the right of individual or collective self-defence recognised by Article 51 of the Charter of the United Nations, will assist the Party or Parties so attacked by taking forthwith, individually and in concert with the other Parties, such action as it deems necessary, including the use of armed force, to restore and maintain the security of the North Atlantic area. Any such armed attack and all measures taken as a result thereof shall immediately be reported to the Security Council. Such measures shall be terminated when the Security Council has taken the measures necessary to restore and maintain international peace and security.<sup>14</sup>



Applied to the cyber domain, the two principles reflect an integrated approach that enables the meeting of what should be a primary interest of the potential targets of a cyberattack: either preventing it in the first place, or, in the event of an actual attack, minimizing its impact. From the perspective of the potential victim, for instance a customer of a large financial interest, this would reflect recognition of the duty to protect and reflect implementing measures to meet that obligation. Much in the same way that governments are obliged to protect their civilian population.

The concept of “customer as partner” in combating cyberterrorism whereby a *triangular relationship* is created between the corporation—the client/customer—law enforcement is far more effective than unnecessarily minimizing—if not denying—the threat.



For a corporation, it is cheaper<sup>15</sup> to react or handle a hack as opposed to spending money on defense and protection. The number of corporations which, in the aftermath of a hack—successful or otherwise—*come forward immediately* and say “we’ve been hacked, we are vulnerable, let’s learn from this” is minimal. That is a lost opportunity for both that company and others. It represents a double victory for the hackers<sup>16</sup>: successful penetration and failure of corporations to learn from each other. While each corporation has interests it must protect, there are sufficient similarities and common values that would facilitate—and welcome—sharing information regarding successful or attempted penetration.

Nevertheless, the reality is that most corporations are extremely hesitant to come forward and acknowledge that they have been hacked. To that end, they are not forthcoming with customers, shareholders, and law enforcement. In addition, they are inhibiting or preventing other corporations from protecting themselves. Perhaps there is a shame element that, despite enormous expenditures on firewalls and IT teams, vulnerability still exists. However, given



the nefariousness of cyberattackers, and the damage caused, it would behoove corporations to put aside that shame factor, and be much more forthcoming.

To that end, it is recommended that failure to share information be deemed a criminal act. The framework provided by the Cybersecurity Information Sharing Act provides the platform for doing so. The reasons for this recommendation are articulated below.

— *Let us consider customers* As a customer of a company that has been hacked, you *immediately* want to know that your privacy is at risk. It is your right to know that “*someone who you did not authorize*” is in possession of your social security number, your health information, and other information of a deeply personal nature. Corporations must have the *immediate* obligation to notify their customers.

— *Let us consider shareholders* Shareholders have significant financial interests at stake. That said, there are significant financial considerations in determining when—and how—to inform shareholders of an attempted or successful cyberattack. Clearly, corporations carefully weigh the impact of a negative response. Nevertheless, corporations must have the absolute responsibility to be as forthcoming as possible to shareholders, and immediately.

— *Let us consider law enforcement* The faster information is provided to law enforcement regarding a cyberattack, the more effectively law enforcement can begin the process of identifying who is responsible. The attacked corporation, ostensibly, has a vested interest in assisting law enforcement; nevertheless, repeated delays in reporting suggest conflict within corporations regardless of ostensible benefits accruing from immediate reporting and information sharing.

To facilitate an institutionalized reporting process we turn our attention to enhanced cooperation among parties directly affected by a cyberattack on a corporation.

### Cooperation to What Extent

A comment and response in the tabletop exercise referenced in the introduction emphasize the complexity of this issue. The dialog can be summarized as follows:<sup>17</sup>

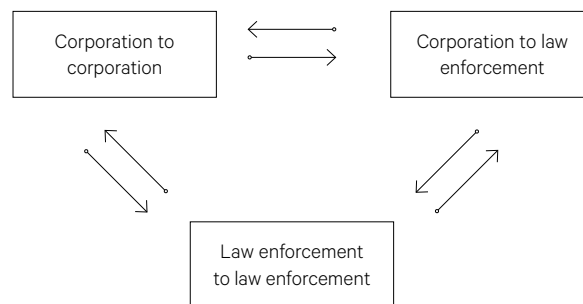
- Amos N. Guiora (ANG): When do you inform customers that you have been hacked, thereby potentially compromising their privacy?
- CEO: Only when the matter is resolved; never while addressing the potential hack.
- Participant/customer: I would want to know as soon as you (the company) know; that way I can take measures to protect myself.
- CEO: My focus will be on resolving the issue rather than informing customers.
- Participant/customer: Meaning you will not enable me to protect my privacy even though you know it is potentially compromised?
- CEO: Correct.
- ANG: Would you inform competitors (in the same industry) that you have been hacked in order to give them “heads up”?
- CEO: No.
- ANG: Why?
- CEO: They’d use it to their benefit, thereby weakening my financial position.
- ANG: Would you inform law enforcement, local or national?
- CEO: That wouldn’t be a priority.





I found the CEO impressive in his candor; it was also clear his company's cyber protocol was sophisticated, reflecting much thought and understanding. That is to his credit. However, I was similarly struck by an unwillingness—well articulated—to cooperate with three distinct actors: customers, competitors, and law enforcement. His answers reflected a “go it alone” approach. This was reinforced when a participant shared with me the following: a chief information security officer (CISO) stated he would share information during an attack only with those corporate entities that he knows and trusts personally.

Business considerations, financial interests, competition, and trade secrets are offered as primary reasons for such an approach. On the other hand, long-term strategic thinking suggests a different approach is essential to successfully countering cyberattacks. That long-term, more strategic approach is primarily predicated on the recognition of a “common enemy” and “combining forces” will significantly enhance the development of more effective countermeasures. The cooperation model suggests cooperation between corporations and between corporations and law enforcement.



Interaction with law-enforcement officials highlighted the lack of cooperation on three distinct levels regarding cybersecurity: corporation-to-corporation, corporation to law enforcement, and law enforcement to law enforcement.

What should corporations do upon discovering that a penetration has occurred?

- conduct an honest assessment of “damage done;”
- take every measure to protect existing data;
- as quickly as possible inform clients and customers;
- immediately inform law enforcement and work hand in hand with them to minimize the damage;
- inform the public;
- cooperate with other corporations, both in order to minimize internal harm and to prevent future attacks.

Undertaking these measures requires sophistication, teamwork, and an ability—and willingness—to analyze internal vulnerabilities, which requires possessing both the necessary tools and competent team members. The willingness, as important as it is, must be complemented by a requisite skill level. Immediate reactivity, which requires integration of teamwork, competence, and willingness, minimizes future harm. However, the primary takeaway from examining how corporations react to successful hacking is a failure to



respond quickly. Whether the failure to respond quickly is deliberate or not is an open question; nevertheless, it demonstrates the challenges in not being able to identify the penetration quickly, and a failure to inform the customer. Both issues—the challenges and informing the client—were commented on by a reader of an earlier draft who noted the following:

Today we are dealing with determined and talented threat actors (hackers), highly skilled and motivated with time and nation-state backing to perform their tradecraft (hack) on a targeted system. Yes, some corporations may have appeared to be easy, but that is the news reporting too. The news is not going to report on a corporation's internal change management, vulnerability management, asset management (hardware/software), IT security processes... these entities are silos/departments that have to come together in order to work a hacking event. And with some recent hacks, the threat actors deliberately obfuscate their activities in order to avoid detection. They are ahead of any system alerting on activity; this is above and beyond virus protection as they portray a normal user on the network, something virus protection does not pay attention to. Whether to inform a customer while an investigation is ongoing, well that is a double-edged sword. Till the full extent of the exposure is determined, advising customers can harm the investigation efforts, especially if you want to see if the threat actor is still active in the system, which is one way of catching him. And the corporation has to determine the full extent of the customer exposure. As we have seen with Equifax [which suffered a data breach in 2017], it had to advise the public over and over about further details of the hack.<sup>18</sup>

Regulatory requirements in the financial industry are now settling on how long you can wait till you advise the public of a hack, and the customers “should” be notified prior to the news/public. The consequences are significant: continued vulnerability; continued threat to customers; potential civil liability in the context of insufficiently protecting customer information/privacy; and liability for failing to notify the customer of the breach and the consequences to the customer.

While focusing on possible lawsuits is understandable, the more important issues are the failure to protect and the failure to inform. The reasons are clear:

- potential customers will hesitate to “bring their business” once they discover failure to protect/failure to inform;
- existing customers may take their business elsewhere if they conclude all reasonable measures were not taken to protect their privacy;
- the broader public will view the corporation negatively in the context of a failure to eliminate cyberattacks and minimize cyber risks BUT the most powerful criticism will be *failure to tell the truth*.

What, then, does that mean for corporations? In stark terms, corporations need to be much more forthcoming. I think there are clear benefits accrued to a corporation in publicly discussing when it has been breached. While the public will express concern in the aftermath of a reported attack, the more long-term reaction will be an appreciation for “speaking the truth.”

Additionally, the knowledge of how a breach occurred, when shared with the public, could prevent a future breach in a different corporation in a similar manner. This change in behavior can have a large positive impact that could potentially affect millions of consumers.

That truth<sup>19</sup> needs to address the following:

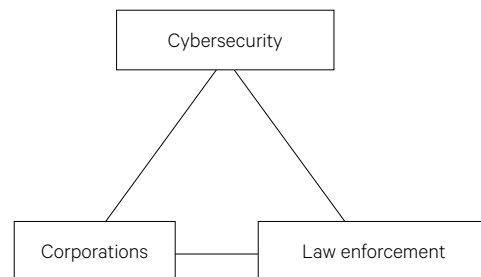
- an acknowledgment of the penetration;
- list of measures undertaken to immediately address the penetration intended to protect customers;
- list of measures intended to protect customers in the future;
- reaching out to other corporations in the context of “information sharing;”
- implement aggressive counter-cyber-measures.



There is risk in this “open” approach; however, from a cost-benefit perspective, the “upside” ultimately outweighs the “downside.” While an element of vulnerability results from openness and candor, an approach emphasizing implementation of measures to prevent future attacks, and honesty with customers and the public, reflects the following: (1) better customer protection; (2) better critical infrastructure protection; and (3) better protection of “larger interest.”

## **Most corporations are extremely hesitant to come forward and acknowledge that they have been hacked. To that end, they are not forthcoming with customers, shareholders, and law enforcement**

Risk mitigation predicated on honesty and proactive aggressive measures is a win-win. From a legal perspective, in terms of minimizing potential impacts of civil suits, a policy of candor and honesty predicated on “we are taking measures to minimize exposure of personal information and learning from it and working hand in hand with customers” will mitigate the possibility of law suits against corporations. Such an approach indicates, from the corporation’s perspective, a willingness to engage different audiences, particularly customers and law enforcement.



For law enforcement to be able to effectively protect corporations, it requires a fundamental change in the context and concept of cooperation and it requires corporations to be more forthcoming to law enforcement.

The following caveat was offered by a reader:

Are you thinking ‘law enforcement’ has the cyber talent to discover vulnerability and how to address it? Who would be the one that could do this: local, state, or federal? The issues when LE [Law Enforcement] is involved: LE is mostly after the ‘who’ committed the hack, whereas the

corporation is focused on getting back into business. Those are competing agendas, one reason LE is not called immediately is that the company is trying to do this work without the distraction of LE. Just because they are LE, does not mean they have the cyber talent readily available to perform an investigation quickly. And depending on the business, the corporation might not have the ability to not be open for business while an investigation is happening.<sup>20</sup>



This cooperation will facilitate law enforcement's understanding of where the hack was, where the specific vulnerability was, and would enhance addressing the points of vulnerability. This can only occur if corporations are much more forthcoming. In that sense, the burden is on them. The failure to work hand in hand with law enforcement prevents development—much less implementation—of a sophisticated, corporate–law enforcement cooperation model. Because of the vulnerability to individuals resulting from a successful cyberattack, there is a pressing need for “out of the box” approaches to law enforcement.<sup>21</sup>

## **Cooperation will facilitate law enforcement's understanding of where the hack was, where the specific vulnerability was, and would also enhance addressing the points of vulnerability**

However, the condition to that approach is the willingness of corporations to view law enforcement as full partners, both preemptively and reactively. To that end, a corporate governance model for cybersecurity is required; while presently untapped, the burden on its development rests with corporations. Law-enforcement officials repeatedly articulated a willingness to closely work with corporations in the development and application of that model, which would emphasize: <sup>22</sup>

- threat identification;
- vulnerability minimization;
- resource prioritization;
- cost-benefit analysis;
- asset protection;
- enhanced understanding of points of vulnerability;
- minimizing impact of future attacks.

So, what do corporations need to do?

Corporate leaders can sit around the table and have endless discussions about points of vulnerability, but the single most effective mechanism to truly understand those points of vulnerability is by conducting sophisticated simulation exercises either in-house or with experts to identify where the corporation is vulnerable.

I would warmly recommend law enforcement have a seat at the table along with other corporations and government officials. Otherwise, the exercise will be akin to an echo chamber, largely ineffective in terms of articulating and implementing an effective cybersecurity policy. I fully understand and respect that for many corporate leaders the idea of institutionalized cooperation with law enforcement, government entities, and other corporations/competitors raises red flags.<sup>23</sup>

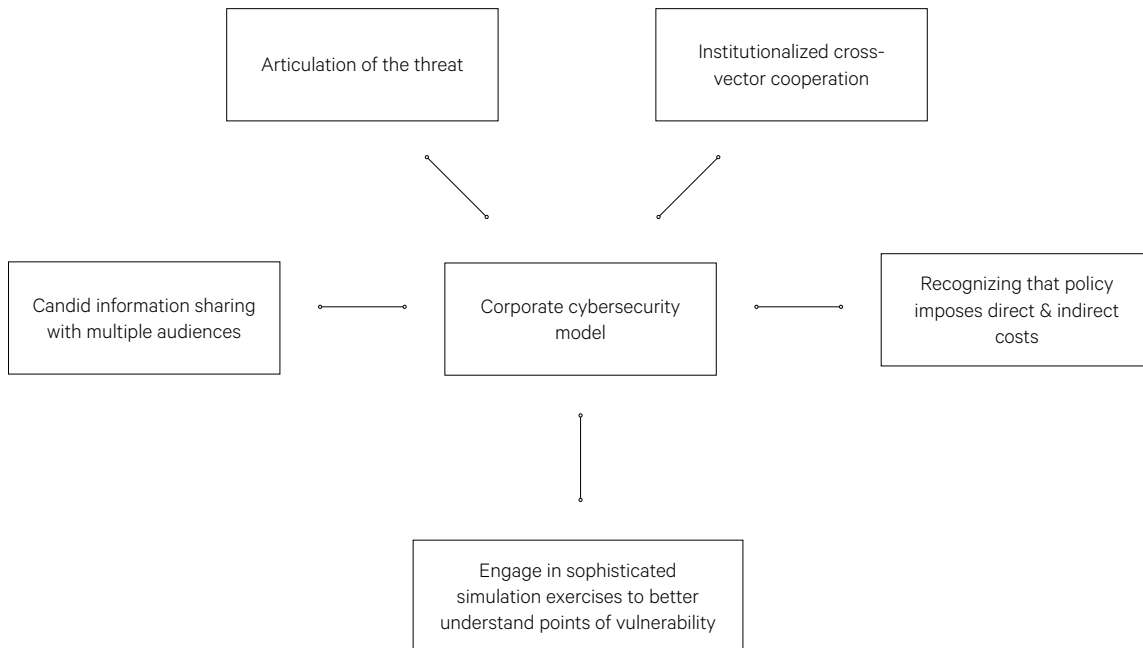




Former managing director of Equifax, Richard Smith, prepares to testify before the United States' Senate's Banking, Housing, and Urban Affairs Committee on October 4, 2017. Smith resigned after it became known that cyber pirates had attacked the credit-standing firm and stolen information on almost 145 million US citizens



However, given the cost, impact, and nefariousness of corporate security hackers, I do not believe that there is any alternative other than to rearticulate the corporate cybersecurity model.



The following are questions to consider for understanding the nature, and extent, of cooperation.<sup>24</sup>

- Should corporations be responsible for their own cybersecurity?
- Should the government force corporations to have a cybersecurity policy?
- Should corporations be required to share relevant cybersecurity information with other corporations, including competitors?
- Should corporations be required to report to law enforcement when they have been attacked?
- Should corporations have a duty to report a cyberattack to shareholders?

As the discussion above illustrates, cooperation is not a “given.” Far from it. However, dismissing it out of hand is too easy. The challenge is to create workable models whereby a proper balance is created that protects the distinct stakeholders without causing unintended financial harm to a cooperating entity. Re-articulated: the challenge, in determining the extent of cooperation, is to convince corporations that cooperation—with customers, law enforcement, and competitors—is advantageous, both tactically and strategically.

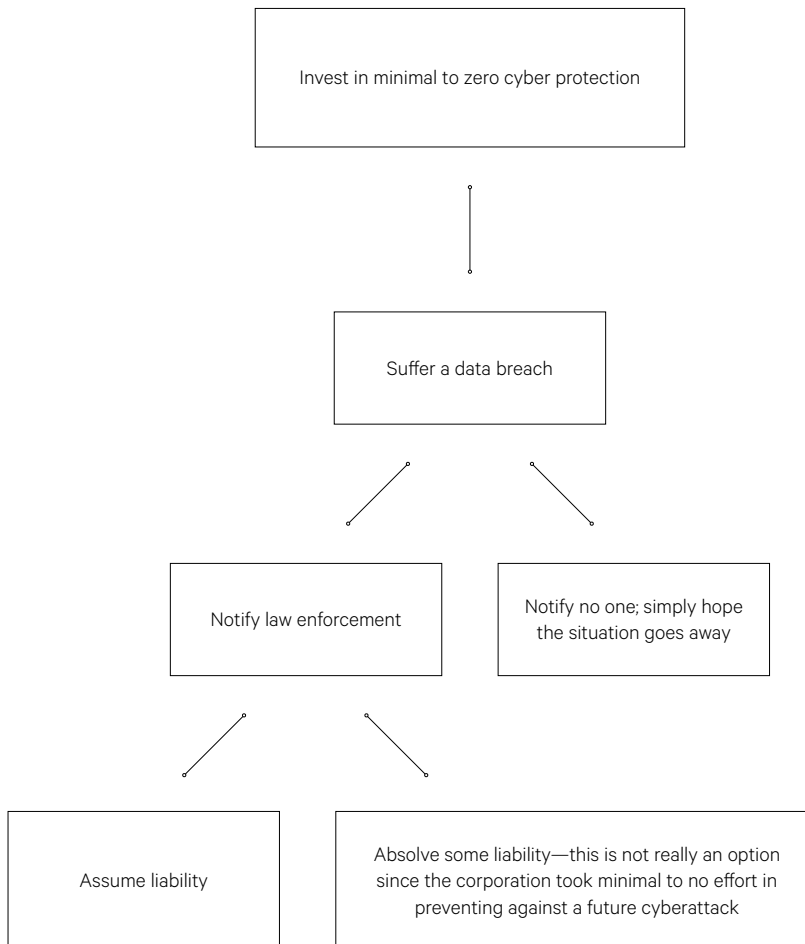
Tactical benefit reflects short-term gain; strategic benefit results in long-term gain. The two are not necessarily synonymous. They can be mutually exclusive. In considering adoption of a cooperation model, the dilemma can be posed as a cost-benefit thought experiment.

To move the ball forward it is essential to consider four different corporations: the two charts below are intended to facilitate understanding of the relationship between corporations and cooperation.



**Data Breach in a Corporation  
(Corporation with Significant Protection)**

**Data Breach in a Corporation  
(Corporation without Significant Protection)**



**Corporation A** Corporation A assumes great responsibility in the amount of information it accesses. Therefore, Corporation A invested significant money and time in cyber protection. It works closely with law enforcement, engages in employee trainings, and actively employs several data experts to protect it against a cyberattack.

Now, despite all its best efforts, Corporation A has been breached. Similar to Target or eBay (which suffered data breaches in 2013 and 2014 respectively), over 100 million customers have now been affected by the breach. The question then becomes, what is the next step, and who has a responsibility in the aftermath of the attack? But, since Corporation A took significant measures to protect itself from a cyberattack, and fell victim anyway, the question becomes should the compensation or retaliation be greater since it took steps to try to prevent the event?

Either way, the first thing that must occur when a company is the victim of a cyberattack is notification to law enforcement. Although companies may not want to report, for fear of customer doubt or repercussion, this must be a legal obligation. Without notification to law enforcement, law enforcement is unable to create patterns or algorithms that could prevent future attacks.

The next thing to consider is whether Corporation A has absolved liability because it took



the necessary precautions and, through no fault of its own, still fell prey to a cyberattack. This is difficult to answer and one that may not be fully answered until additional, relevant legislation is put into place.

**Corporation B** Corporation B, like Corporation A, is one of the largest corporations in America. From that, Corporation B assumes great responsibility in the amount of information it accesses. However, Corporation B has not invested significant money or time in cyber protection. Rather, its Board of Directors, which is actively aware of the threat of cybersecurity, voted to delay any financial or personnel investment in the pursuit of cyber protection because it is expensive, and the corporation is in the business of making money. This issue is the pinnacle of the protection versus profit debate.

Now, imagine Corporation B has been breached.

Similar to Target or eBay, over 100 million customers have now been affected by the breach. The question then becomes, what is the next step, and who has a responsibility in the aftermath of the attack? As mentioned earlier, the United States government may take it upon itself to get involved when corporations exist of a certain size, as it has done previously. But the question becomes, since Corporation B did not take significant measures to protect itself from a cyberattack, and fell victim, should the compensation or retaliation be less since it did not take steps to prevent the event?

## **Law enforcement should have a seat at the table along with other corporations and government officials. Otherwise, the exercise will be akin to an echo chamber, largely ineffective in terms of articulating and implementing an effective cybersecurity policy**

As mentioned before, either way, the first thing that has to occur when a company is the victim of a cyberattack is notification of such an attack to law enforcement. Usually, most companies do not want to report, whether from fear of customer doubt or shareholder perception, or other repercussions; this needs to be a legal obligation. The next thing to consider is whether Corporation B is increasingly liable for its negligence.

Unlike Corporation A, its liability cannot be absolved in any way because it did not take necessary precautions. Despite likely recommendations from the Chief Intelligence Officer, or other employees in the corporation, Corporation B chose profit over protection, and became an easy target. The question then becomes, due to its negligence, should the compensation or protection post-attack be less? That is a question that will need to be determined by future legislation.

**Corporation C** Corporation C, unlike Corporations A and B, is one of the smaller corporations in America, a small-town business held closely by a few family members. From that, Corporation C assumes significantly less responsibility in the amount of information it accesses.

However, like Corporation A, Corporation C has invested significant money and time in cyber protection. It works closely with law enforcement, engages in employee training, and actively employs a data expert to protect itself against a cyberattack.

Despite Corporation C's best efforts, it has been breached. However, unlike Corporations A and B, the breach does not affect over 100 million customers. Rather, the breach affects simply 5,000 individuals. The question still becomes, what is the next step, and who has





a responsibility in the aftermath of the attack? It is safe to assume that the United States government is less likely to get involved when the breach is so minimal, as compared to the Corporation A and B scenario.

However, the question to be asked is, since Corporation C took significant measures to protect itself from a cyberattack, and fell victim anyway, should it be compensated or supported in some way, more so than a company that made no effort to protect itself against a cyberattack?

Corporation C, despite its small size in comparison to Corporations A and B, must still have a legal obligation to report to law enforcement when it is the victim of a cyberattack. After which, the next question to consider is whether Corporation C absolves liability because it took the necessary precaution and, through no fault of its own, still fell prey to a cyberattack.

**Corporation D** Corporation D, unlike Corporations A and B, but similar to Corporation C, is one of the smaller corporations in America, a small-town business held closely by a few family members. From that, Corporation D assumes significantly less responsibility in the amount of information it accesses. However, like Corporation B, Corporation D has not invested significant money or time in cyber protection.

Rather, the corporation decided to delay any financial or personnel investment in the pursuit of cyber protection because it is expensive, and the corporation is focused on making money. This issue, as demonstrated with Corporation B as well, is the pinnacle of the protection versus profit debate.

Due to Corporation D's lack of effort, it has been breached. However, unlike Corporations A and B, the breach does not affect over 100 million customers. Rather, the breach affects simply 5,000 individuals. The reasoning and considerations now become very similar to the questions that occurred with the breach in Corporation C. The question still becomes, what is the next step, and who has a responsibility in the aftermath of the attack?

It is safe to assume that the United States government is less likely to get involved when the breach is so minimal, as compared to Corporations A and B. But the question becomes, since Corporation D did not take significant measures to protect itself from a cyberattack, and fell victim, should the compensation or retaliation be less since it did not take steps to prevent the event?

Corporation D, despite its small size in comparison to Corporations A and B, must still have a legal obligation to report when it is the victim of a cyberattack. The next question to consider is liability. Unlike Corporations A and C, its liability cannot be absolved in any way because it did not take necessary precautions.

Despite likely recommendations from the Chief Intelligence Officer, or other employees in the corporation, Corporation D chose profit over protection, and became an easy target. The question then becomes, due to its negligence, should the compensation or protection post-attack be less? That is a question that will need to be determined by future legislation.

## The Good, the Bad, and the Ugly of Cooperation

Easier said than done.

That oft-repeated maxim applies to the corporation model that is the underlying premise of this article. Corporation is not cost-free. There is understandable concern regarding vulnerability, exposure, and unwarranted risk, perhaps with minimal "upside."



The CEO referenced in the section “Cooperation to What Extent” was eloquent in his conciseness; cooperation, for him, is an absolute nonstarter. I heard a similar theme from other corporate leaders. To my surprise, a similar refrain was articulated by federal law-enforcement officials when asked whether they would cooperate with local enforcement. The answer was brief, and embarrassing: “We don’t cooperate with local law enforcement.” Embarrassing because the brief, actually brusque, answer was given to local law-enforcement officials at a meeting in my presence. However, in its brevity the official highlighted an important—albeit troubling—concern: the extent to which law enforcement cooperates among itself, much less with external entities.

This struck me then, and continues to this day to reflect short-term arrogance, putting one’s own “turf” ahead of the public good. In a recent meeting with federal law-enforcement officials, I was assured such a response would be met with great disfavor by senior officials. I hope that is the case.

Assurances aside, it is important that cooperation—among local, state, and federal officials—be institutionalized, not left to the whims and fancies of particular officials. That seems to serve the public interest and public good which is the primary obligation of law enforcement. Institutionalizing cooperation—among the various stakeholders—requires time, dedication, and resources. It is not a “given” for actors representing different cultures, norms, and mores to immediately—and instinctively—have protocols in place easily facilitating cooperation. This will require a concerted effort.

## **Assurances aside, it is important that cooperation—among local, state, and federal officials—be institutionalized, not left to the whims and fancies of particular officials. That seems to serve the public interest and public good which is the primary obligation of law enforcement**

Resources must be allocated for such implementation; common language must be found; common goals need to be identified; understandable hesitation, reticence, and skepticism must be held in abeyance. This requires a dedicated effort by all involved; the motivation for doing so must focus on the larger good, rather than temporary costs and burdens. Otherwise, the undertaking falls flat.

Conversely, the positives of cooperation far outweigh the conflicts and tensions. This is viewed from the perspective of public safety. That, frankly, must be the most important consideration in weighing-up the adoption of a cooperation model.

### **Final Word**

The theme of cooperation has been the centerpiece of this article. That is not by chance.

I served as legal advisor to a congressional task force under the auspices of the House Committee on Homeland Security addressing US Homeland Security policy. The principle of cooperation was uppermost in my mind in the development—and implementation—of effective cyber countersecurity. The discussion in this paper is intended to facilitate the reader’s understanding of the need to develop institutionalized cooperation and, simultaneous to that, recognizing the difficulty in such an effort.



Former United States National Security Agency surveillance center in Bad Aibling, south of Munich. In June 2013, Edward Snowden's revelations of massive surveillance programs by the NSA thwarted diplomatic relations between the United States and its European partners







The reasons are varied; whether financially driven, as is the case with corporations, or “turf” and budget, as was explained to me by law-enforcement officials, the consequences are clearly predictable. The beneficiary of a consistent lack of cooperation is the wrongdoer; the victims are plentiful. The lack of cooperation theme was consistently expressed to me while writing my book. It was repeated when conducting the tabletop exercise. In all three instances—the US Congress, book research, exercise—those with whom I engaged were uniform regarding cooperation, whether proactive or reactive: the notion does not fit their “model.” The mantra was repeated a sufficient number of times.

This was in direct contrast to a meeting I held in Israel with a leading cyber expert. The conversation on that occasion was extraordinarily insightful, shedding light on the intersection between national security and cybersecurity. More importantly, it highlighted the crucial role government can—and should—play with respect to cyber. Our conversation focused on questions of law and policy; the technical matters, while undoubtedly important, were not at the forefront of what we discussed. What particularly impressed me—in the context of cooperation—was the enormous benefit accrued when public and private sectors joined forces and cooperated.

That is not intended as a gloss over inevitable tensions, jealousies, and competition between the two. It was, however, in sharp contrast to the discussions I had with US law-enforcement officials. The difference between the two approaches was jarring. The consequences are obvious. It is for that reason that the theme of cooperation occupies such importance in my book: *Cybersecurity: Geopolitics, Law, and Policy*.

## The principle of cooperation should prevail in the development and implementation of effective cyber countersecurity

Many of us have been victims of harm, whether on a personal, professional, or community basis. Our vulnerability to cybercrime is well documented; there is no need to repeat the litany of incidents, ranging from the irritating to the truly catastrophic.

It is clear that individuals in groups, worldwide, are dedicated to continuously seeking ways in which to use cyber to their advantage and to our disadvantage. There is, truly, an “us-them” with respect to cyber. The harms posed by cybercriminals and cyberterrorists are significant; of greater concern are the future harms they are, undoubtedly, planning to impose on society. Of that, I have no doubt.

There is a profound lack of consensus regarding the question of government involvement. Perhaps as a direct reflection of my background in the Israel Defense Forces, I am—frankly—baffled—by the hesitation repeatedly expressed to me regarding the role of government in cyber protection. I believe that cyberattacks need to be perceived as similar to physical attacks.

The consequence of that, for me, is clear: an attack on an American corporation warrants government response. While that requires the cooperation discussed above, the benefits—short and long term alike—significantly outweigh any negative consequences regarding government “over” involvement. Frankly, the stakes are too high to resort to tired clichés and irrelevant mantras regarding privacy concerns. That is not to minimize the question of privacy—NSA leaks disturbingly highlight the reality of government intrusion—but it is to suggest that cyber threats require a balanced and nuanced approach. Summarily dismissing government involvement is shortsighted and ultimately counterproductive.



Hand in hand with government involvement is the question of self-defense and collective self-defense. In actuality, the two are directly related and cannot be separated one from the other.

Self-defense is a critical question in the cyber discussion. The inquiry is whether the nation state owes a duty to corporations and individuals who have been victimized by a cyberattack. It is not an abstract question, but one rather intended as a concrete query.

There are great risks in imposing “response” burdens on the nation state in the aftermath of a cyberattack. This is possible only if the cooperation model is adopted. If all potential—and actual—targets of a cyberattack are to be “stuck” in a noncooperation paradigm, then the forces of darkness are the obvious winners. That is clear.

Nevertheless, enough opposition has been consistently articulated—by the private and public sector alike—that it must give pause. That opposition, and certainly skepticism, are legitimate and cannot be dismissed with a casual wave of the hand. And yet.

And yet, the persistent threat posed by potential cyberattacks, not to mention the actual impact in the aftermath of a successful attack, warrants careful consideration of a proposal to legislatively mandate cooperation requirements and obligations. Given the consequences of a successful attack, cooperation among the different actors outlined in this paper would have the potential to minimize the consequences of a particular impact.

At the very least, cooperation would contain the consequences by ensuring that other impacted entities—yes, competitors—would be positioned to implement, proactively, protection measures while simultaneously ensuring that law enforcement—national and local—be fully engaged in minimizing the fallout.

There are different mechanisms for determining the viability of such an undertaking. As was made clear to me in the tabletop exercise, communication and dialog are excellent starting points. To bring the different actors together, to explore how to respond and—hopefully—minimize the fallout of an attack is an important step. It is recommended that such exercises include a wide range of participants, including elected officials, law enforcement (local, state, and national), corporate leaders (large and small; private and public), municipal officials, representatives of consumer advocacy groups, and members of the public.

Such an undertaking will enable the community—defined broadly and narrowly—to determine what is the best mechanism to ensure cooperation in the face of a cyberattack.

That is a discussion we must have. Postponement in exploring cooperation models is far from “cost-free”; quite the opposite: it comes with great cost. The existing measures provide an effective platform from which to build. The challenge to do so is great. The time to do so is now.

## Acknowledgments

I owe many thanks to the following individuals who graciously read through previous drafts and made insightful comments and suggestions that significantly strengthened the final product; needless to say, all mistakes, errors, and omissions are exclusively my responsibility: Anne-Marie Cotton, Senior Lecturer in Communication Management, Artevelde University College, Ghent (Belgium); Jessie Dyer JD (expected 2019), S. J. Quinney College of Law, University of Utah; Brent Grover, President, Brent Grover & Co.; Professor Daniel Shoemaker, University of Detroit Mercy; Judy A. Towers, Head of Cyber Threat Intelligence, SunTrust Bank.

## Notes

1. A reader of an earlier draft suggested this statement applies to 99% of business executives with whom he has interacted.
2. Amos N. Guiora, *Cybersecurity: Geopolitics, Law, and Policy*, Routledge, 2017.
3. Due to "Chatham House" rules, the identities of participants and their comments may not be noted; the exercise was conducted under the auspices of the National Cyber Partnership.
4. For a thoughtful discussion regarding the Social Contract, please see <https://www.iep.utm.edu/soc-cont/>.
5. <https://www.investopedia.com/terms/s/sarbanesoxleyact.asp>.
6. <https://www.cisecurity.org/newsletter/cybersecurity-information-sharing-act-of-2015/>.
7. <https://www.cisecurity.org/newsletter/cybersecurity-information-sharing-act-of-2015/>.
8. <https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive>.
9. <https://www.us-cert.gov/ais>.
10. Domestic issues: not only stealing ID but also property, as illustrated in this commercial: [https://www.youtube.com/watch?v=\\_CQA3X-qNgA](https://www.youtube.com/watch?v=_CQA3X-qNgA).
11. A reader of an earlier draft noted: "A lot of people would say that cybersecurity is nothing more than risk management"; e-mail in my records.
12. Illustrated by the Belgian agency: <https://www.youtube.com/watch?v=jop2l5u2F3U>.
13. <http://www.un.org/en/sections/un-charter/chapter-viii/>.
14. [https://www.nato.int/cps/ic/natohq/official\\_texts\\_17120.htm](https://www.nato.int/cps/ic/natohq/official_texts_17120.htm).
15. A reader of a previous draft noted the following points: based on which evidence? What about negative impact on reputation? Loss of confidence? Intangibles compared to financial figures of costs to redesign the "firewalls"... however, brand equity has become an important element too.
16. A reader of a previous draft noted: wonderfully developed by Andrew Ehrenberg (London School of Economics) in his double jeopardy theory applied to

big companies versus small ones: the small ones are twice "victim" of their size: although they invest, the big are investing more = they have to invest proportionally X more to get a smaller result...

17. In accordance with Chatham House rules.
18. E-mail exchange with reader; e-mail in author records.
19. See articles by Finn Frandsen, Timothy Coombs, Robert Heath, Winni Johansen on this theme, for example: "Why a concern for apologia and crisis communication?" Frandsen and Johansen have just published their latest theoretical developments in: *Organizational Crisis Communication: A Multivocal Approach*.
20. E-mail exchange with reader; e-mail in author records.
21. The following caveat was offered by a reader of an earlier draft: "Besides awareness—this also takes a lot of time... We are only in the foothills of producing the talent and people in LE rarely have the bandwidth to incorporate another skill set"; e-mail in my records.
22. A reader of a previous draft noted: this is a good statement, but I have not seen LE be business savvy enough to perform the articulation of the items in the list. This is where industry regulators step in with the knowledge to provide these items, not LE. And they are advising DHS as they are setting the same standards in the list.
23. A reader of a previous draft noted that this is being done as we speak in the US financial sector; and the businesses were candid in their "whys" of remediation processes while all of us that represented regulators/LE were waiting to be told. Also for a company to determine a hack has occurred instead of just a system glitch can take days to determine, especially if the threat actor has modified system logs and deleted their code/actions/steps. Making things look like normal system issues is a sophisticated threat actor. Hoping this ability does not become easy for an "everyday Joe" threat actor.
24. A reader of a previous draft noted that these questions are now being answered by industry regulators, that is, New York state implemented these requirements for any businesses headquartered in NY by the end of 2018; these businesses will start to be audited by the state to ensure cybersecurity has been addressed. Again, do not forget what happens when companies work toward compliance only: if the compliance standards do not stay current, then a company can satisfy regulators/compliance but still not be cybersecure. Most companies that have been hacked today are regulatory compliant, the regulator just did not stay current. This is an evolving door scenario. Companies will be compliant to the minimum as to go further costs money.

## **Edition**

BBVA

## **Project direction and coordination**

Chairman's Advisory, BBVA

## **Texts**

Michelle Baddeley

Joanna J. Bryson

Nancy H. Chau

Barry Eichengreen

Francisco González

Amos N. Guiora

Peter Kalmus

Ravi Kanbur

Ramón López de Mántaras

María Martínón-Torres

José M. Mato

Diana Owen

Alex Pentland

Carlo Ratti

Martin Rees

Victoria Robinson

Daniela Rus

José Manuel Sánchez Ron

Vivien A. Schmidt

Samuel H. Sternberg

Sandip Tiwari

Yang Xu

Ernesto Zedillo Ponce de León

## **Edition and production**

Turner

## **Publishing coordination and graphic edition**

Nuria Martínez Deaño

## **Gráphic design and layout**

underbau

## **Translation**

Wade Matthews

## **Copyediting**

Harriet Graham

## **Printing**

Artes Gráficas Palermo

## **Binding**

Ramos

## **Digital publishing**

Comando g

## **Images**

Daniel Acker/Bloomberg via Getty Images: p. 248; Jason Alden/Bloomberg via Getty Images: p. 457; AMA/Corbis via Getty Images: p. 346 (image above); American Cancer Society/Getty Images: p. 212; Christopher Anderson/Magnum Photos: p. 406 (image above); Marco Ansaloni/SCIENCE PHOTO LIBRARY: p. 86 (image above); Francisco Asencio: p. 107; Patrick Aventurier/Getty Images: p. 228; Gonzalo Azumendi/Getty Images: p. 95; Matthias Balk/Picture Alliance via Getty Images: p. 198; Joseph Barrak/AFP/Getty Images: p. 446 (image above); Juan Barreto / AFP/Getty Images: p. 303; BBVA: p. 31; Jonas Bendiksen/Magnum Photos: pp. 428 (image below) and 430; Noah Berger/AFP/Getty Images: p. 379; Kena Betancur/AFP/Getty Images: p. 322; Jabin Botsford/The Washington Post

via Getty Images: p. 382 (image above); Brill/ullstein bild via Getty Images: pp. 220 (image below) and 222; James Brittain/View Pictures/UIG via Getty Images: p. 136; Frederic J. Brown/AFP/Getty Images: p. 413; Michael Christopher Brown/Magnum Photos: pp. 446 (image below) and 448; SIP/UIG via Getty Images: p. 227; Martin Bureau/AFP/Getty Images: p. 12 (image above); Andrew Burton/Getty Images: p. 401; José Calvo/SCIENCE PHOTO LIBRARY: p. 211; John B. Carnett/Bonnier Corporation via Getty Images: p. 193; CERN, 2017-2018: pp. 56 (image below) and 58; CERN, 2005-2018: p. 62; CERN, 2008, a beneficio de CMS Collaboration/Fotografía de Michael Hoch y Maximilien Brice: pp. 66-67; Creapole/Alex Pentland: p. 107; Pigi Cipelli/Archivio Pigi Cipelli/Mondadori Portfolio via Getty Images: pp. 126 (image below) and 128; Khaled Desouki/AFP/Getty Images: p. 428 (image above); Al Drago/Bloomberg via Getty Images: p. 288 (image above); Patrick T. Fallon/Bloomberg via Getty Images: p. 220 (image above); Gregor Fischer/Picture Alliance via Getty Images: p. 253; J. Emilio Flores/Corbis via Getty Images: pp. 288 (image below) and 290; Frederick Florin/AFP/Getty Images: pp. 168-169; Katherine Frey/The Washington Post via Getty Images: p. 139; Christopher Goodney/Bloomberg via Getty Images: pp. 346 (image below) and 348; Fritz Goro/The LIFE Picture Collection/Getty Images: p. 126 (image above); Louisa Gouliamaki/AFP/Getty Images: p. 332; Steve Gschmeissner/SCIENCE PHOTO LIBRARY: p. 244; Noriko Hayashi/Bloomberg via Getty Images: p. 361; Ann Hermes/The Christian Science Monitor via Getty Images: p. 235; Stefan Heunis/AFP/Getty Images: p. 439; Zakir Hossain Chowdhury/Barcroft Media via Getty Images: p. 364 (image above); Lewis Houghton/SCIENCE PHOTO LIBRARY: p. 229; Institute for Stem Cell Research via Getty Images: p. 186 (image above); Wolfgang Kaehler/LightRocket via Getty Images: pp. 274-275; Michael Kappeler/Picture Alliance via Getty Images: p. 21; Manjunath Kiran/AFP/Getty Images: pp. 24-25; Kyodo News via Getty Images: pp. 194-195; Wolfgang Kumm/Picture Alliance via Getty Images: p. 216; Andrew Lichtenstein/Corbis via Getty Images: p. 336; Matthew Lloyd/Bloomberg via Getty Images: p. 132; Majority World/UIG via Getty Images: p. 258 (image above); Marek Mis/SCIENCE PHOTO LIBRARY: p. 204 (image above); MIT Media Lab/Getty Images: p. 230; Jeff J. Mitchell/Getty Images: p. 389; John Moore/Getty Images: p. 102 (image above); Eduardo Munoz Álvarez/Getty Images: pp. 406 (image below) and 408; Samuel Nacar/SOPA Images/LightRocket via Getty Images: pp. 356-357; NASA: p. 34 (image above); NASA, ESA and the Hubble sm4 ERO Team: p. 49; NASA/Dimitri Gerondidakis: pp. 44-45; NASA/MSFC/David Higginbotham: pp. 34 (image below) and 36; NASA/JPL-Caltech/MSSS: p. 41; NASA/Science Photo Library: p. 75;

Mandel Ngan/AFP/Getty Images: p. 456; Robert Nickelsberg/Getty Images: pp. 418-419; Leonard Ortiz/Digital First Media/Orange County Register via Getty Images: pp. 238 (image below) and 240; Punit Paranjpe/AFP/Getty Images: p. 370; Paolo Pellegrin/Magnum Photos: p. 115; Gilles Peress/Magnum Photos: pp. 12 (image below), 14, 326 (image below) and 328; Spencer Platt/Getty Images: p. 339; Pool/ Tim Graham Picture Library/Getty Images: p. 353; Balint Pornecezi/AFP/Getty Images: pp. 382 (image below) and 384; Sergey Pyatakov/Sputnik/SCIENCE PHOTO LIBRARY: p. 137; Aamir Qureshi/AFP/Getty Images: p. 310 (image above); REUTERS/Iván Alvarado: pp. 310 (image below) and 312; REUTERS/Rebecca Cook: p. 337; REUTERS/Kevin Coombs: p. 18; REUTERS/Jonathan Ernst: pp. 280 and 298; REUTERS/Robert Galbraith: p. 156; REUTERS/Kim Hong-Ji TPX: pp. 364 (image below) and 366; REUTERS/Adam Hunger: p. 374; REUTERS/Aaron Josefczyk: p. 144; REUTERS/Kham: p. 294; REUTERS/Adrees Latif: pp. 258 (image below) and 260; REUTERS/Jason Lee: p. 299; REUTERS/David Loh: pp. 86 (image below) and 88; REUTERS/Jon Nazca: pp. 318-319; REUTERS/Thomas Peter: p. 150 (image above); REUTERS/Michaela Rehle: p. 472; REUTERS/Joshua Roberts: p. 213; REUTERS/Pascal Rossignol: pp. 392-393; REUTERS/Axel Schmidt: p. 265; REUTERS/Gregory Scruggs: p. 26; REUTERS/Tyrone Siu: p. 249; REUTERS/Shannon Stapleton: pp. 150 (image below) and 152; Reuters/Stringer: p. 342; REUTERS/Benoit Tessier/Pool: pp. 204 (image below) and 206; REUTERS/Paulo Whitaker: p. 238 (image above); Bertrand Rindoff Petroff/Getty Images: p. 114; Janine Schmitz/Photothek via Getty Images: p. 119; Roberto Schmidt/AFP/Getty Images: p. 434; Georgy Shafeev/SCIENCE PHOTO LIBRARY: p. 83; Luke Sharrett/Bloomberg via Getty Images: p. 163; Qilai Shen/Bloomberg via Getty Images: p. 283; Prakash Singh/AFP/Getty Images: p. 423; Jeff Spicer/Getty Images: pp. 102 (image below) and 104; Melanie Stetson Freeman/The Christian Science Monitor via Getty Images: p. 172; Newsha Tavakolian/Magnum Photos: p. 396; Javier Trueba/MSF/SCIENCE PHOTO LIBRARY: p. 98; Peter Tuffy, University of Edinburg/Science Photo Library: p. 56 (image above); UN Photo/Eskinder Debebe: p. 371; VCG/VCG via Getty Images: pp. 122 and 326 (image above); Rudolf Viced/Getty Images: p. 94; Mark Wilson/Getty Images: p. 465; Hitoshi Yamada/NurPhoto via Getty Images: p. 438; Beata Zawrzel/NurPhoto via Getty Images: pp. 186 (image below) and 188.

© of the publication, BBVA, 2018

© of the texts, their respective authors, 2018

© of the translation, his author, 2018





## ACCESS THE FULL BOOK IN SPANISH

- ¿Hacia una nueva Ilustración? Una década trascendente

### HOW TO CITE THIS BOOK

Towards a New Enlightenment? A Transcendent Decade: Madrid, BBVA, 2018.

## ALL THE OPENMIND COLLECTION TITLES

