

Chapter 3: CRYSTAL BALL AND MAGIC WAND¹ (Version 0.1) –

The dangerous promise of Big Data

by Dirk Helbing

Every activity in the world now leaves ever more digital traces. What if a "wise king" or "benevolent dictator" had real-time access to all the data in the world? Could he take perfect decisions to benefit society? Could he predict the future? Could he control the further development of the world? What would be the limits and side effects of such an approach? Or is the attempt to create a Crystal Ball to predict the world and a Magic Wand to control it a dangerous dream?

Data sets bigger than the largest library

The idea that we could solve the problems in the world, if we just had enough data about them, is intriguing. In fact, we are now entering an era of "Big Data" – masses of information, mostly in digital form, about all aspects of our lives, institutions and cultures. It will probably not be long before each newborn baby will have its genome sequenced at birth. Every purchase we make on the Internet releases data about our location, preferences, and

¹ Dear Reader,

thank you for your interest in this chapter, which is thought to stimulate debate.

What you are seeing here is work in progress, a chapter of a book on the emerging Digital Society I am currently writing. My plan was to elaborate and polish this further, before I share this with anybody else. However, I often feel that it is more important to share my thoughts with the public now than trying to perfect the book first while keeping my analysis and insights for myself in times requiring new ideas.

So, please apologize if this does not look 100% ready. Updates will follow. Your critical thoughts and constructive feedback are very welcome. You can reach me via dhelbing@ethz.ch or @dirkhelbing at twitter.

I hope these materials can serve as a stepping stone towards mastering the challenges ahead of us and towards developing an open and participatory information infrastructure for the Digital Society of the 21st century that would enable everyone to take better informed decisions and more effective actions.

I believe that our society is heading towards a tipping point, and that this creates the opportunity for a better future. But it will take many of us to work it out. Let's do this together!

Thank you very much, I wish you an enjoyable reading,

Dirk Helbing

PS: Special thanks go to the FuturICT community and to Philip Ball.

finances that will be stored somewhere and quite possibly used without our consent. Cell phones disclose where we are, and private messages and conversations are not really private at all. Books are being digitized beyond the advent of printing, and are available in immense, searchable databases of words that are now being mined in "culturomics" studies that put history, society, art and cultural trends under the lens. Aggregated data can be used to reveal unexpected facts, such as flu epidemics being inferred from Google searches.

This avalanche of data is ever increasing: with the introduction of technologies such as Google Glass, people will have the option of documenting and archiving almost every aspect of their lives. Big Data such as credit-card transactions, communication and mobility data, public news, Google Earth imagery, comments and blogs, are creating an increasingly accurate digital picture of our physical and social world, including all its social and economic activities.

"Big Data" will change our world. The term, coined more than 15 years ago, means data sets so big that one can no longer cope with them with standard computational methods. To benefit from Big Data, we must learn to "drill" and "refine" data, i.e. to transform them into useful information and knowledge. The global data volume doubles every 12 months. Therefore, each year we produce as much data as in all previous years together together.

These tremendous amounts of data relate to four important technological innovations: the *Internet*, which enables our global communication, the *World Wide Web* (WWW), a network of globally accessible websites that evolved after the invention of hypertext protocol (HTTP), the emergence of *Social Media* such as Facebook, Google+, Whatsup, or Twitter, which have created social communication networks, and the emergence of the "*Internet of Things*" (IoT), which allows sensors, smartphones, gadgets, and machines ("things")

to connect to the Internet. Note that there are already more things connected to the Internet than humans.

Meanwhile, the data sets collected by companies such as ebay, Walmart or Facebook, reach the size of petabytes (1 million billion bytes) – one hundred times the information content of the largest library in the world: the U.S. Library of Congress. The mining of Big Data opens up entirely new possibilities for the optimization of processes, the identification of interdependencies, and the support of decisions. However, Big Data also comes with new challenges, which are often characterized by four criteria: *volume* (the file sizes and number of records are huge), *velocity* (the data evaluation has often to be done in real-time), *variety* (the data are often very heterogeneous and unstructured), and *veracity* (the data are probably incomplete, not representative, and contain errors).

Gold rush for the 21st century's oil

When the social media portal WhatsApp with its 450 million users was recently sold to Facebook for \$19 billion – almost half a billion dollars was made per employee. There's no doubt that Big Data create tremendous business opportunities – not just because of its value for, say, marketing, but because the information itself is becoming monetarized.

Technology gurus preach that Big Data is becoming the new oil of the 21st century: a commodity that can be tapped for profit. With the virtual currency *BitCoin* temporarily becoming more valuable than gold, one can even literally say that data can be turned into value to an extent we only knew from fairy tales. Even though many sets of Big Data are proprietary, the consultancy company McKinsey recently estimated the potential value of Open Data alone to be 3 to 5 trillion dollars per year. If the worth of this publicly available information were to be evenly apportioned among the public itself, it would bring \$500 to each person in the world.

The potential of Big Data spans all areas of social activity: from natural language processing to financial asset management, or to a smart management of cities that better balances energy consumption and production. It could enable better protection of our environment, risk detection and reduction, and the discovery of opportunities that would otherwise be missed. And it could make it possible to tailor medicine to patients, thereby increasing drug effectiveness, accelerating drug discovery and reducing side effects.

Big Data applications are now spreading very rapidly. They enable personalized services and products, open up entirely new possibilities to optimize production and distribution processes or services, allow us to run "smart cities," and reveal unexpected interconnections between our activities. Big Data also hold great potential for fostering evidence-based decision-making, particularly in big business and politics. Where is all of this leading us?

In this chapter I explain what Big Data can and cannot do. In particular, I will show that it will never be enough on its own to avoid crises and catastrophes, or to solve all societal problems. Indeed, the suggestion sometimes heard – that Big Data is the key to the future – can be misleading in pretty dangerous ways. Most obviously, it could precipitate a descent into an authoritarian surveillance state where there is very little personal liberty or autonomy, and where no one can have any more secrets. But even if that were not to happen, Big Data could create a false sense that we can control our own destiny, if only we have enough data. Information is potentially useful, but it can only release its potential if it is coupled to a sound understanding of how complex social systems work.

Big Data fueling super-governments

The development of human civilization has depended on the establishment of mechanisms that promote cooperation and social order. One of these is

based on the idea that everything we do is seen and judged by God. Bad deeds will be punished, while good ones will be rewarded. The age of information has inspired the dream that we might be able to know and control everything ourselves: to acquire God-like omniscience and omnipotence. There are now hopes and fears that such power lies within the reach of huge IT companies such as Google or Facebook and clandestine secret services such as the CIA or NSA. CIA Chief Technology Officer Ira "Gus" Hunt² has explained how easy it is for such institutions to gather a great deal of information about each of us:

"You're already a walking sensor platform... You are aware of the fact that somebody can know where you are at all times because you carry a mobile device, even if that mobile device is turned off. You know this, I hope? Yes? Well, you should... Since you can't connect dots you don't have, it drives us into a mode of, we fundamentally try to collect everything and hang on to it forever... It is really very nearly within our grasp to be able to compute on all human generated information."

Could such a massive data-collection process be good for the world, helping to eliminate terrorism and crime? Or should we fear that the use of this information will undermine human rights and the basis of democratic societies and free economies? In this chapter, I explore the possibilities and limits of such an approach, for better or worse.

Imagine you are the president of a country, intending to ensure the welfare of all its people. What would you do? You might well wish to prevent wars and financial crises, economic recessions, crime, terror, and the spread of diseases. You may want people to be rich, happy, and healthy. You would like to avoid unhealthy drug consumption, corruption, and perhaps traffic jams as

² See: <http://www.businessinsider.com/cia-presentation-on-big-data-2013-3?op=1> and <http://gigaom.com/2013/03/20/even-the-cia-is-struggling-to-deal-with-the-volume-of-real-time-social-data/2/>. For similar recent FBI priorities see http://www.slate.com/blogs/future_tense/2013/03/26/andrew_weissmann_fbi_wants_real_time_gmail_dropbox_spying_power.html

well. You would like to ensure a safe, reliable supply of food, water, and energy, and to keep the environment in good shape. In sum, you would like to create a prosperous, sustainable and resilient society.

What would it take to achieve all this? You would certainly have to take the right decisions and avoid those that would have harmful, unintended side effects. So you should know about alternatives for impending decisions, along with their opportunities and risks. For your country to thrive, you would have to avoid ideological, instinctive or traditional decisions in favor of evidence-based decisions. To have the evidence needed to inform this decision-making, you would need a lot of data about all quantifiable aspects of society, and excellent data analysts to interpret it. You might well decide to collect all the data you can get, just in case it might turn out to be useful one day to counter threats and crises, or to exploit opportunities that might arise.

Previously, rulers and governments were not in this position: they generally lacked the quality or quantity of data needed to take well-informed decisions. But that is now changing. Over the past several decades, the processing power of computers has exploded, roughly doubling every 18 months. The capacity for data storage is growing even faster. The amount of data is doubling every year. With the now emerging "Internet of Things," cell phones, computers and factories will be connected to the most mundane devices – coffee machines, fridges, shoes and clothes – creating an overwhelming stream of information that feeds an ocean of "Big Data."

Humans governed by computers?

The more that data is generated, stored and interpreted, the easier is it to find out about each individual's behavior. Everyone's computer and everyone's device-encoded behavior (such as the record of our movements produced by the cell phones we carry) leaves a unique fingerprint, such that it is possible to know our interests, our thinking, our passions, and our feelings. Some

companies analyze "consumer genes" to offer personalized products and services. They have already collected up to 3,000 personal data from almost a billion people in the world: names, contact data, incomes, consumer habits, health information, and more. This is pretty much everyone with a certain level of income and Internet connectivity.

Would it be beneficial if a well-intentioned government had access to all this data? It could help politicians and administrations to take better informed decisions: to reduce terrorism and crime, say, and to use energy more efficiently, protect our environment, improve traffic flows, avoid financial meltdowns, mitigate recessions, enhance our health system and education, and provide better fitting services to citizens.

Moreover, could a government use the information not only to understand but also to predict our behavior, and map out the course of our society? Could it optimize our social systems and take the best decisions for everyone?

In the past, we have used supercomputers for almost everything except understanding our economy, society, and politics. Every new car or airplane is designed, simulated, and tested on a computer. Increasingly, so are new drugs. Thus, why shouldn't we use computers to understand and guide our economy and society too? In fact, we are slowly moving towards that very situation. As a minor (yet revealing) example, since their early days computers have been used for traffic control. Today's economic production and the management of supply chains would not be conceivable without computer control as well, and large airplanes are now controlled by a majority decision among several computers. Computers can already beat the best chess players, and about 70 percent of financial transactions are executed by trading computers in the meantime. IBM's Watson computer has started to take care of some customer hotlines, and computer-driven Google cars will soon move around without a driver, perhaps picking up the goods we ordered on the Web without us being present. In all these cases, computers already

do a better job than humans. Why shouldn't they eventually make better policemen, administrators, lawyers, and politicians?

It no longer seems unreasonable, then, to imagine a gigantic computer program that could simulate the actions and interactions of all the humans in the world, perhaps even equipping these billions of agents with cognitive abilities and intelligence. If we fed these agents with our own personal data, would they behave as we do? In other words, would it be possible to create a virtual mirror world? And would machine learning eventually be able to build the computer agents so similar to us that they would take decisions undistinguishable from ours? Attempts to construct or at least envisage such a scheme are already underway. If they were realized, would they represent a kind of Crystal Ball with which we could predict the future of society?

The prospect might sound unnerving to some, but in principle the potential benefits aren't hard to see. There are many huge problems that such a predictive capability might help to solve. The financial crisis has created global losses of at least 20 trillion US dollars. Crime and corruption consume about 2-5% of the gross domestic product (GDP) of all nations on earth – about 2 trillion US dollars each year. The lost output of the US economy as a result of the 9/11 terror attacks is estimated to be of the order of 90 billion dollars. A major influenza pandemic infecting 1% of the world's population would cause losses of 1-2 trillion dollars per year. Cybercrime costs Europe alone 750 billion Euros a year. The negative economic impact of traffic congestion amounts to 7-8 billion British Pounds in the United Kingdom alone.

If a computer simulation of the entire global socio-economic system could produce just a 1 percent improvement in dealing with these problems, the benefits to society would be immense. And in fact, if experiences with managing smaller complex social systems this way are any guide, an improvement of 10-30 percent seems conceivable. Overall this would amount to savings of more than 1 trillion dollars annually. Even if we had to invest

billions in creating such a system, the benefits could exceed the investments hundred-fold. Even if the success rates were significantly smaller, this would represent a substantial gain. It would be hard to see how any responsible politician could decline to support such an investment.

But would such a system work as one might hope? Is Big Data all a government needs to get our world under control?

Crystal Ball and Magic Wand

Recent studies using smartphone data and GPS traces suggest that more than 90 percent of the mobility of people – where they will be at a certain time – can be forecast, because of its repetitive nature. If other aspects of our behavior show the same degree of predictability, it's not hard to imagine that the trajectory of society can indeed be mapped out in advance, with all that this entails for successful social planning. While some people might not like this prospect at all, many would perhaps appreciate a predictable life.

How far does this idea extend? If we have enough data about every aspect of life, could we become omniscient about the future? In order to achieve that, we would need to be able to manipulate people's choices using the information provided to them. Personalized Internet searches, systems such as *Google Now*, and personalized advertising are already going in this direction.

But given the overwhelming amount of data available, it needs to be filtered before to be useful. Such filtering will inevitably be done in the interests of those who do the filtering. For example, companies want potential customers to see their ads and buy their products. The better people's characteristics are known, the easier it becomes to manipulate their choices. A recent, controversial *Facebook* experiment with 600 million users showed that it is possible to manipulate people's feelings and mood. Therefore, it's not hard to

imagine that omniscience might indeed imply omnipotence: those who know everything could control everything. Let's call the hypothetical tool creating such power a "Magic Wand".

Assuming that we had a Magic Wand, could we take the right decisions for our society, or even for every individual? Many people might say that forecasting societal trends is different from forecasting the weather: the weather does not care about the forecast, but people will respond to it, and this will defeat the prophecy. That seems to imply that successful forecasting of societal developments would require that people don't know about the forecasts, while governments do. This again suggests that one would need a secretly operating authority advising the government about the right decisions, and that it would use the Magic Wand according to the evidence provided by the data-collecting Crystal Ball. Could such a scheme work?

A New World Order based on information?

Our "wise king" or "benevolent dictator" would probably see the Crystal Ball and the Magic Wand as perfect tools to create social order. Singapore is sometimes seen as an approximation of such a system. The country has indeed been enormously successful in the past decades, but despite great advances and fast economic growth, people's satisfaction has decreased. Why?

A wise king would certainly sometimes have to interfere with our individual freedoms, if we would otherwise take decisions that would create more damage for the economy and society than benefits. This might end up in a situation where we would always have to execute what the government wants us to do, pretty much as if they were commands from God. If we were manipulated in our decision-making, this might even happen without our knowledge. Although the wise king would not be able to fulfill our wishes all the time, on average he might create better outcomes for everyone as long as

do as we are told. Sure, this sounds dystopian, but let us nevertheless pursue the concept for a while to see whether it is feasible in principle. If we obediently followed the dictates of the wise king, could this improve the state of the world and turn it into a perfectly working clockwork?

Why top-down control is destined to fail

In short, it would not work. This kind of top-down management, even if guided by comprehensive information, is destined to fail. This book is, therefore, concerned with elaborating alternative and better ways of using data, which are compatible with constitutional rights and cultural values such as privacy. But let us first figure out what are the reasons why a well-working Crystal Ball and Magic Wand can't exist.

One of the problems is statistical in nature. To distinguish “good” from “bad” behavior, we need criteria that clearly separate the two. In general, however, reliable criteria of this sort don't exist. We face the problems of false positive classifications (false alarms, so-called type I errors) and false negatives (type II errors, where the alarm is not triggered when it should be).

For example, imagine a population of 500 million people, among which there are 500 terrorists. Let's assume that we can identify terrorists with an extremely impressive 99 percent accuracy. Then there are 1 percent false negatives (type II error), which means that 5 terrorists are not detected, while 495 will be discovered. It has been revealed that about 50 terror acts were prevented over the past 12 years or so, while a few, such as the one during the Boston marathon, were not prevented even though the terrorists were listed in some databases of suspects (in other words, they turned out to be false negatives).

How many false positives (false alarms) would the above numbers create? If the type I error is just 1 out of 10,000, there will be 50,000 wrong suspects, while if it is 1 in one thousand then there will be 500,000 wrong suspects. If it is 1 percent (which is entirely plausible), there will be 5 million false suspects!

It has been reported that there are indeed between 1 and 8 million people on lists of suspects in the USA. If these figures are correct, this would mean that for every genuine terrorist, up to 10,000 innocent citizens would be wrongly categorized as potential terrorists. Since the 9/11 attacks, about 40,000 suspects have had to undergo special questioning and screening procedures at international airports, even though in 99 percent of these cases it was concluded that the suspects were innocent. And yet the effort needed to reach even this level of accuracy is considerable and costly: according to media reports, it involved around a million people who had a National Security Agency (NSA) clearance on the level of Edward Snowden.

So, large-scale surveillance is not an effective means of fighting terrorism. This conclusion has, in fact, been reached by several independent empirical studies. Applying surveillance to the whole population is not sensible, for the same reasons why it is not generally useful to apply prevention-oriented medical tests or medical treatments to the entire population: since such mass screenings imply large numbers of false positives, millions of people might be wrongly treated, often with negative side effects on their health. Thus, for most diseases, patients should be tested only if they show worrying symptoms.

Besides these errors of first and second kind, one may face errors of third kind, namely inappropriate models for separating "good" from "bad" cases. For example, unsuitable risk models have been identified as one reason for the recent financial and economic crisis. The risks of many financial products turned out to be wrongly rated, creating immense losses. Adair Turner, head of the UK Financial Service Authority, has said that there is

“a strong belief ... that bad or rather over-simplistic and overconfident economics helped create the crisis. There was a dominant conventional wisdom that markets were always rational and self-equilibrating, that market completion by itself could ensure economic efficiency and stability, and that financial innovation and increased trading activity were therefore axiomatically beneficial.”

Limitations of the Crystal Ball

One might think that errors of first, second, and third kind could be overcome if only we had enough data. But is this true? There are a number of fundamental scientific factors that will impair the Crystal Ball's functioning (see Information Box 1). The problem known as "Laplace's Demon" reflects on the history-dependence of future developments, and our inability to ever measure all the past information needed to predict the future, even if the world changed according to deterministic rules (that is, if there were no randomness). This is why we are still influenced by cultural inventions, ideas, and social norms that are hundreds or thousands of years old.

Furthermore, turbulence and chaos are well-known properties of many complex dynamical systems. These factors imply that even the slightest change in the system at a certain point in time may fundamentally change the outcome over a sufficiently long period of time. The phenomenon, also named the "butterfly effect," is well-known to impose limits on the time horizon of weather forecasts.³ In social systems as in the weather system, this extreme sensitivity to small but unpredictable disturbances arises from the complexity of the system: the existence of many interdependencies between the component parts.

Furthermore, we can determine the parameters of our models only with a finite accuracy. However, even small changes in these parameters may fundamentally change the outcome of the model. There is also a problem of ambiguity: the same information may have several different meanings depending on the respective context, and the particular interpretation we choose may influence the future course of the system. Beyond this, we also know that certain statements are fundamentally "undecidable" in the sense that there are questions that cannot be answered with formal logic. Lastly, too much information may reduce the quality of predictions because of over-fitting,

³ These prediction limits are not just a matter of getting enough measurement data and having a sufficiently powerful computer – one cannot get beyond a certain precision because of the physical nature of the underlying process.

spurious correlations, and herding effects. The Information Box at the end of this chapter elaborate these points in more detail.

So one can say that Big Data is not the universal tool that it is often claimed to be.⁴ Any attempt to predict the future will be limited to probabilistic and mostly short-term forecasts. It is therefore dangerous to suggest that a Crystal Ball could be built that would reliably predict the future.

Limitations of the Magic Wand

If the Crystal Ball is cloudy, it doesn't augur well for the Magic Wand that would depend on it. In fact, top-down control is still very ineffective, as the abundance of problems in our world shows. To control complex systems, i.e. to force them to behave in certain ways, we often do not understand them well enough and lack effective means. Therefore, in many cases attempting to control a complex dynamical system in a top down way undermines its functionality. The result is often a broken system, for example, an accident or crisis.

An example of the failure of top-down control is the fact that even the most sophisticated technological control mechanisms for airplane flight safety increased it less efficiently than introducing a non-hierarchical culture of collaboration in the cockpit, when co-pilots were encouraged to question the decisions and actions of the pilot. In another example, the official report on the Fukushima nuclear disaster in Japan stresses that it was not primarily the earthquake and tsunami that were responsible for the nuclear meltdowns, but

“our reflexive obedience; our reluctance to question authority; our devotion to ‘sticking with the program’; our groupism.”

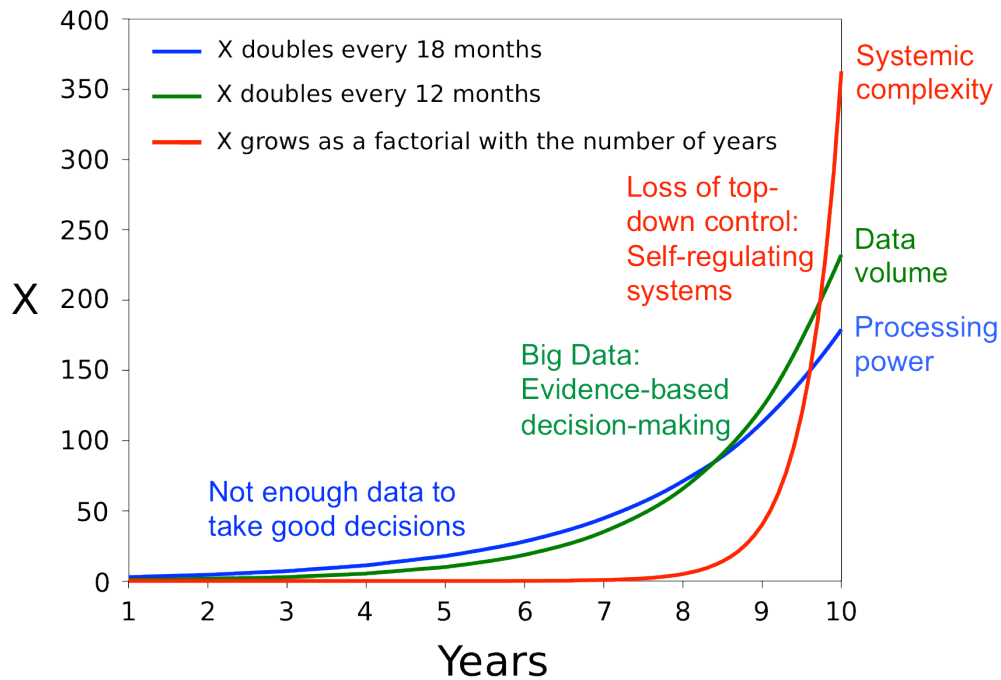
In other words, the problem was too much top-down control. Attempts to

⁴ To convince me of the opposite, in analogy to the "Turing test" checking whether a computer can communicate undistinguishable from a human, one would have to demonstrate that a computer system passes the "Helbing test," i.e. finds all fundamental laws of physics discovered by scientists so far, just by mining the experimental data accumulated over time.

control complex systems in a top-down way are also very expensive, and we find it increasingly hard to pay for them: most industrialized countries already have debt levels of at least 100 or 200 percent of their gross domestic product. But do we have any alternatives? In fact, the next chapters of this book will elaborate one.

Complexity is the greatest challenge, but also the greatest opportunity

There are further reasons why the concept of a "super-government", "wise king" or "benevolent dictator" can't really work. These are related to the complexity of socio-economic systems. There are at least four kinds of complexity that matter: dynamic complexity, structural complexity, functional complexity and algorithmic complexity. The problem of complex dynamics has been addressed in the previous chapter. Here, I will focus on implications of structural, functional and algorithmic complexity. In fact, with a centralized super-computing approach we can only solve those optimization problems, which have sufficiently low algorithmic complexity. However, many problems are "NP-hard," i.e. so computationally demanding that they cannot be handled in real-time even by super-computing. This problem is particularly acute in systems that are characterized by a large variability. In such cases, top-down control cannot reach optimal results. In the next chapter, I will illustrate this by the example of traffic light control.



Given the quick increase in computing power, couldn't we overcome this challenge in the future? The surprising answer is "no." While the processing power doubles every 18 months (blue curve in the illustration above), the amount of data doubles every year (green curve above). This implies that we are heading from a situation in which we did not have enough data to take good decisions, to a situation where we can take evidence-based decisions. However, despite the rising processing power, we will be able to process a decreasing share of all the data existing in the world. Moreover, the lack of processing power will be quickly increasing. So we are moving to a situation where we can shed light on everything with a spotlight, but many things will remain unseen in the dark. This creates a new kind of problem: paying too much attention to some problems, while neglecting others. In fact, governments didn't see the financial crisis coming, they didn't see the Arab Spring coming, they didn't see the crisis in the Ukraine coming, and they didn't see the Islamic State (IS) fighters in Iraq coming. Thus, keeping a well-balanced overview of everything will become progressively more difficult. Instead, politics will be increasingly driven by problems that suddenly happen to gain public attention, i.e. made in a reactive rather than anticipatory way.

But let's now have a look at the question of how the world is expected to change depending on its complexity. The possibility to network the components of our world creates ever more options. We have, in fact, a combinatorial number of possibilities to produce new systems and functionalities. If we have two kinds of objects, we can combine them to produce a third one. These three kinds of objects allow us to create six ones, and those already 720. This is mathematically reflected by a factorial function, which grows much faster than exponentially (see the red curve above). For example, we will soon have more devices communicating with the Internet than people. In about 10 years from now, 150 billion (!) things will be part of the Internet, forming the "Internet of Things." Thus, even if we realize just every thousandth or millionth of all combinatorial possibilities, the factorial curve will eventually overtake the exponential curves representing data volumes and computational power. It has probably overtaken both curves already some time ago.

In other words, attempts to optimize systems in a top-down way will become less and less effective – and cannot be done in real time. Paradoxically, as economic diversification and cultural evolution progress, a "big government", "super-government" or "benevolent dictator" would increasingly struggle to take good decisions, as it becomes more difficult to satisfy the diverse local expectations and demands. This means that centralized governance and top-down control are destined to fail. Given the situation in Afghanistan and Iraq, Syria, Ukraine, and the states experiencing the Arab Spring, given the financial, economic and public debt crisis, and given the quick spreading of the Ebola disease in Africa, have we perhaps lost control already? Are we fighting a hopeless battle against complexity?

Simplifying our world by homogenization and standardization would not fix the problem, as I will elaborate in the chapter on the Innovation Accelerator. It would undermine cultural evolution and innovation, thereby causing a failure to adjust to our ever-changing world. Thus, do we have alternatives? Actually, yes: rather than fighting the properties of complex systems, we can use them

for us, if we learn to understand their nature. The fact that the complexity of our world has surpassed our capacity to grasp it, even with all the computers and information systems assisting us, does not mean that our world must end in chaos. While our current system is based on administration, planning, and optimization, our future world will be built on evolutionary principles and collective intelligence, i.e. intelligence surpassing that of the brightest people and best expert systems.

How to get there? In the next chapters, I will show how the choice of suitable local interaction mechanisms can, in fact, create desirable outcomes.

Information and communication systems will enable us to let things happen in a favorable way. This is the path we should take, because we don't have better alternatives. The proposed approach will create more efficient socio-economic institutions and new opportunities for everyone: politics, business, science, and citizens alike. As a positive side effect, our society will become more resilient to the future challenges and shocks that we will surely face.

Conclusions

"Big Data" has great potential, in particular for better, evidence-based decision-making. But it is not a universal solution, as it is often suggested. In particular, data-driven approaches are notoriously bad at predicting systemic shifts, where the entire way of organizing or doing things change. Moreover, like any technology, Big Data can be seriously misused, posing a "dual use problem" (see the Information Box 2 below). Without suitable precautions – for example, the use of "data safes," decentralization, encryption, the logging of large-scale data-mining activities, the limitation of large processing volumes to qualified and responsible users, the accountability of Big Data users for damage created by them, and large fines in cases of damage, misuse, or discrimination – mining Big Data may create massive problems (intentionally or not). It is, therefore, crucial to design socio-technical systems in ways that promote their ethical use.

INFORMATION BOX 1: Limitations to Building a Crystal Ball

Sensitivity - When all the data in the world can't help

How close can computer-modeled behavior ever come to real human social behavior? To specify the parameters and starting conditions of a computer model, these are varied by calibration procedures until the difference between measurement data and model predictions becomes as small as possible. However, the best fitting model parameters are usually not the correct parameters. These parameters are typically within a certain "confidence interval." But if the parameters are randomly picked from the confidence interval, the model predictions may vary a lot. This problem is known as sensitivity.

Turbulence and chaos

Two further problems of somewhat similar nature are "chaos" and "turbulence." Rapid flows of gases or liquids produce swirly patterns – the characteristic forms of turbulence. In chaotically behaving systems, too, the motion becomes unpredictable after a certain time period. Even though the way a "deterministically chaotic" system evolves can be precisely stated in mathematical terms, without random elements, the slightest change in the starting conditions can eventually cause a completely different global state of the system. In such a case, no matter how accurately we measure the initial conditions of the system, we will effectively not be able to predict the later behavior.

Ambiguity

Information can have different meanings. In many cases, the correct interpretation can be found only with some additional piece of information: the context. This contextualization is often difficult and not always available when needed. Different pieces of information can also be inconsistent, without any means of resolving the conflict.

A typical problem in "data mining" challenges is that data might be plentiful, but inconsistent, incomplete, and not even representative. Moreover, a lot of it might be wrong, because of measurement errors, misinterpretations or application of wrong procedures, or manipulation.

Laplace's Demon and measurement problems

Laplace's Demon is a hypothetical being who could calculate all future states of the world, if he knew the exact positions and speeds of all particles and the physical laws governing their motion and interactions. Laplace's Demon cannot exist in reality, not least because of the fundamental limitation that measurements to determine all particle speeds would be impossible due to the restriction of special relativity: all velocities must be less than the speed of light. This would prevent one from gathering all the necessary data.

Information overload

Having a lot of data does not necessarily mean that we'll see the world more accurately. A typical problem is that of "over-fitting," where a model characterized by many parameters is fitted to the fine details of a data set in ways that are actually not meaningful. In such a case, a model with less parameters might provide better predictions. Spurious correlations are a somewhat similar problem: we tend to see patterns where they actually don't exist (see <http://www.tylervigen.com/> for some examples).

Note that we are currently moving from a situation where we had too little data about the world to a situation where we have too much. It's like moving from darkness, where we can't see enough, to a world flooded with light, in which we are blinded. We will need "digital sunglasses": information filters that will extract the relevant information for us. But as the gap between the data that exists and the data we can analyze increases, it might become harder to pay attention to those things that really matter. Although computer processing power doubles every 18 months, we will be able to process an ever decreasing fraction of all the data we possess, because the data storage capacity doubles every year. In other words, there will be increasing volumes of data that will never be touched.

Herding

When people feel insecure, they tend to follow decisions and actions of others. This produces undesirable herding effects. The economics Nobel laureates George Akerlof (*1940) and Robert Shiller (*1946) have called this behaviour "animal spirits," but in fact the idea of herding in economics goes back at least to the French mathematician Louis Bachelier (1870-1946). Bubbles and crashes in stock markets are examples of where herding can lead.

Randomness and innovation

Randomness is a ubiquitous feature of socio-economic systems. However, even though we would often like to reduce the risks it generates, we would be unwise to try to eliminate randomness completely. It is an important driver of creativity and innovation; predictability excludes positive surprises and cultural evolution. We will see later that some important and useful social mechanisms can only evolve in the presence of randomness. Although newly emerging behaviors are often costly in the beginning, when they are in a minority position, the random coincidence or accumulation of such behaviors in the

same neighborhood can be very beneficial, and such behavior may then eventually succeed and spread.

INFORMATION BOX 2: Side effects of massive data collection

Like any technology, Big Data has not only great potential but also harmful side effects. Not all Big Data applications come with these problems, but they are not uncommon. What we need to identify, are those problems that can lead to major crises rather than just localized, small-scale defects.

Crime

In the past years, cybercrime has increased exponentially, costing Europe alone around 750 million EUR per year. Some of this has resulted from the undermining of security standards (for example those of financial transactions) for the purpose of surveillance. Other common problems are data theft or identity theft, data manipulation, and the fabrication of false evidence. These crimes are often committed by means of “Trojan horses”, computer codes that can steal passwords and PIN codes. Further problems are caused by computer viruses or worms that damage software or data.

Military risks

Because most of our critical infrastructures are now connected with other systems via information and communications networks, they have become pretty vulnerable to cyber attacks. In principle, malicious intruders can manipulate the production of chemicals, energy (including nuclear power stations), and communication and financial networks. Attacks are sometimes possible even if the computers controlling such critical infrastructures are not connected to the Internet. Given our dependence on electricity, information and money flows as well as other goods and services, this makes our societies vulnerable as never before. Coordinated cyber-attacks could be

launched within microseconds and bring the functioning of our economy and societies to a halt.

The US government apparently reserves the right to respond to cyberwar with a nuclear counter-strike. We are now seeing a digital arms race for the most powerful information-based surveillance and manipulation technologies. It is doubtful whether governments will be able to prevent serious misuse of such powerful tools. Just imagine, a Crystal Ball or Magic Wand or other powerful digital tools would exist. Then, of course, everyone wanted to use them, including our enemies, and criminals as well. It is obvious that, sooner or later, these powerful tools would get into wrong hands and finally out of control. If we don't take suitable precautions, mining massive data may (intentionally or not) create problems of any scale – including digital weapons of mass destruction. Therefore, international efforts towards confidence-building and digital disarmament are crucial and urgent.

Economic risks

Cybercrime poses obvious risks to the economy, as do illicit access to sensitive business secrets and theft of intellectual property. Loss of customer trust in products can cause sales losses of the order of billions of dollars for some companies. Systems that would not work effectively without a sufficient level of trust include electronic banking, sensitive communication by email, eBusiness, eVoting, and social networking. Yet more than two thirds of all Germans say they do not trust government authorities and Big Data companies any longer to not misuse their personal data. More than 50 percent even feel threatened by the Internet. The success of the digital economy is further threatened by information pollution, for example, spam and undesired ads.

Social and societal risks

To contain "societal diseases" such as terrorism and organized crime, it often seems that surveillance is needed. However, the effectiveness of mass surveillance in improving the level of security is questionable and hotly debated: the evidence is missing or weak. At the same time, mass surveillance erodes the trustful relationship between citizens and the state. The perceived loss of privacy is also likely to promote conformism and to endanger diversity and useful criticism. Independent judgments and decision-making could be undermined. Excessive state control of the behavior of citizens would, therefore, impair our society's ability to innovate and adapt.

For such reasons, the constitutions of many countries consider it of fundamental importance to protect privacy, informational self-determination, private communication, and the principle of assumed innocence without proof of guilt. These things are also considered to be essential for human dignity, and elementary preconditions for democracies to function well.

However, today the Internet lacks good mechanisms for forgetting, forgiveness, and re-integration. There are also concerns that the increasing use of Big Data could lead to greater discrimination, which in turn could promote increasing fragmentation of our society into subcultures. For example, it is believed that the spreading of social media has promoted the polarization of US society.

Political risks

It is often pointed out that leaking confidential political communication can undermine the success of sensitive negotiations. Moreover, if incumbent governments have better access to Big Data applications than parties in opposition, this could result in unfair competition and non-representative election outcomes. Last but not least, in the hands of extremist political groups or criminals, Big Data could become a dangerous tool for acquiring and exerting power.