Alessandro Vespignani
and Rosita Copioli

The Oracle and the Algorithm

Book proposal

Why do humans love rules so much? The answer might actually be very simple: because rules makes us feel in control. They allow us to know what to expect from reality, what we need to tackle it, and what consequences could follow. The most sophisticated form of control is «prediction» – man's inherent ability to anticipate and investigate the future, and exorcise our fears and our doubts by understanding what will happen next. Isn't that the reason we check the weather forecast? Staying ahead of weather – while planning a meeting, a dinner party, a getaway, or simply because the weather outside often affects our mood – allows us to address what might happen without risking setbacks. If all functions well – and it usually does – we make no mistakes in planning our vacation. (It goes without saying that every year those same forecasts save thousands of live from tropical storms, typhoons and catastrophic events in which forecasting can be used to call for evacuations, alerts, and weather updates.)

Isn't that the same feeling that made some Greek warriors retreat into the cave of the Cumaen Sybil to make her predict their fate in the coming battle? Weather websites – even though they differ greatly – offer a very similar experience to the palm leaves the Sybil used to write on.

Science has always considered prediction to be a fundamental concept. Indeed, it has spoiled us, trying to overprotect us like a mother does. Do we want to know when the next moon eclipse will take place? Easy-peasy. We can just go online, and find out about every lunar eclipse in the coming centuries –with clear-cut precision. Somehow predictions have become a commodity: through scientific rules and prediction, we are able to devise new materials and study their properties on a computer. For example, we could conceptualize the aerodynamic properties of a plane before assembling even the smallest part of it. These are extraordinary successes that have allowed humanity to take giant leaps. On the other hand, this doesn't reflect an increased ability to control a whole other array of events – which, incidentally, are those that most concern us.

Can we govern conflicts and wars? Predict epidemics, or the management of humanitarian emergencies?

As of now, it would seem that we cannot: we even have trouble predicting something as simple as a gridlock near a traffic light. We live in an increasingly interconnected world, which is increasingly exposed to small perturbations that can reverberate throughout entire systems – sometimes leading to catastrophic consequences.

Why can't we predict and control these? Because at the heart of all of this is *us* – the social atom. Unfortunately, there's a rule that applies to the social atom, something that for a long time made us think that we were unpredictable, unforecastable. Every one of us has a cognitive, psychological complexity that makes us different from physical atoms. However, according to some schools of thought, we have similar properties as physical atoms. The advent of sociology introduced a social systems physics that is based on the social atom as individual, as well as its molecules, such as family and workplace. By assembling these molecules, we obtain a social aggregate. Once we consider the social aggregate as an emergent phenomenon – as claimed by sociologists – then we are no longer so unpredictable: in fact, 90 percent of the things we do are easily foreseeable.

This ambitious vision had to tackle several hardships. But two great revolutions took place in the last decades that allowed this vision to come closer and closer to reality. First was the revolution established by «complexity science», which deals with every kind of system – physical as well as biological – consisting of many interacting individuals. It is an extraordinary science that taught us that the queen bee myth does not exist: there is no queen bee to rule her worker bees. Instead, society can accomplish great things without any leader, or explicit project and simply through emergent phenomena and interactions. Since our human society is similar to a hive, this applies to us as well.

Then another revolution took place: a data revolution. To truly study the social atom, we must consider the data that describe it. We live in a data-driven age. Every year, the amount of socio-economic data that we gather is more than all the data from our past. The data collected in one week far outweigh those that were collected during the entire Roman Empire. Data come in various forms, not just from the digitalization of archives. They come from widespread technologies, or from the fact that everyone could potentially wear small sensors – like geolocators – whereby their interactions could be very accurately measured. Scientists could examine millions of us interacting, like teachers who watch children during classes or at recess. They could observe how we speak to each other, how we behave on a physical level.

Because of these revolutions, we can fine tune history, as we are able to observe society from increasingly greater distances and look into data that produced using smartphones, computers, and other devices. We can follow our movements as we board a plane or hop on a train, and within hours we are in a different city, even a different continent: we can see *everything*.
But these are simply data, so neither *prediction* nor control. To get to prediction one must build up «models» capable of putting the data together, of analyzing them. Models coincide with an understanding of the figures that allow us to project reality into the future: they are our crystal ball, our flock of birds. To use

them, though, we must leave behind linear thinking and try something new: «computational» thinking. If we truly have no compass, we can get our bearings with a computer. A computer is no longer a tool dealing exclusively with numbers, or performing numerical operations. It has become a true laboratory, a place where data can be analyzed, where statistical regularities can be deducted, where models can be created, and where we will finally be able to understand and explore the world at large. It's the «macroscope» that allow us to look into the collective sphere, the very large scale. At the same time, it will be based on the aggregation of many small elements, allowing us to understand what those elements are doing and –more importantly – what they are going to do.

The prediction of epidemics, civil wars, gunfights, fatal accidents, anomalous migrations: these are just some of the circumstances that could be predicted using a computer, in the same way that we currently forecast next weekend's weather. To transform the unpredictable into rules: this a revolutionary, ambitious challenge.

Synopsis

1. From the natural world to societal physics

For three centuries, natural science allowed us to succeed in understanding and predicting developments in the world surrounding us. From tides to meteorology, from planetary orbits to the subatomic world, and cellular world through the mapping of human genome. It has given us a reassuring sense of being in control, but that clashes with our insecurity over our inability to apply the same method to other fields. This is the case with financial crises, political instability, and the spreading of epidemics. Man –defined here as the social atom – stands at the center of these phenomena, which don't play by the strict rules of the physical universe. Society is a tangled web of economical, political, cultural and physical interactions and dependencies that seem to refute all attempts of prediction. This chapter will talk about how the science of complex systems succeeded in outlining a series of mathematical rules and general principles by discrediting the queen bee, or leader myth, and focusing on dynamic interactions among the system elements. These rules and principles can lead the way toward a conceptual understanding and modeling of phenomena such as social contagion, conflicts and emergencies.

2. The Science of Prediction

In the past decade the science of prediction has been redefined by the big data revolution. A growing number of socioeconomic data has been swiftly made available thanks to the digitalization of our world. But that's not all: the advent of the mobile phone and other widespread technologies – including the web and a multitude of social networks – is producing an unprecedented deluge of information regarding our daily life: how we get from one place to the next, what we talk about, how do we get our information. These data have become a veritable fuel to the predictive algorithms currently pervading our lives. In 2008 Chris Andersen wrote an article for *Wired* where he bluntly said that we were entering a world where a humongous quantity of data and algorithms would substitute any other approach. In this chapter, we describe how automatic learning and artificial intelligence—instead of being explicitly programmed—deployed this revolution, allowing the algorithms to gain knowledge by assimilating bulk data.

3. How many copies will this book sell?

Nowadays automatic learning methods can interpret x-ray tests and optimize online advertising. But above all, it can provide forecasts. Algorithms have become the digital fortune tellers we rely on for investigating the future. This chapter will deal with the deployment of algorithms to guess our political inclinations, our taste in movies, and even whether or not we are on the brink of divorce. It will also deal with how predictions can go further, outlining the success of a new novel or how far the career of a scientist or an artist will go (and we will of course apply these instruments to try and forecast how many copies this book will sell in various international markets). Lastly, we will talk about their applications in medicine – a field where a fresh wave of predictive and analyzing algorithms capable to establish the individual risk of contracting a disease is transforming how we think about prevention and treatment.

4. The parable of Google Flu: traps and limits of big data

In February 2013, Google Flu Trend (GFT) made the news. Unfortunately, not for the reasons that Google managers and the creators of this flu tracking system would have hoped for. GFT foresaw more than two times the number of medical examinations due to Influenza-like illnesses (ILI), when compared to the Centers for Disease Control and Prevention (CDC) in the USA, which base their estimates on surveillance reports issued by American laboratories. What can we learn from their mistake? Where do algorithms misfire and what are the limits of prediction? We answer these questions as well as show the perils of algorithms due to unconscious prejudices either implemented in the software by coders, or learned by artificial intelligence on the basis of previously analyzed data.

5. Synthetic worlds

Man can succeed where machines fail. In these cases, the algorithm absorbs theoretical knowledge, turning it into a calculus of sorts, an expression of the system's operating mechanisms that one wants to predict. This is exactly what we do daily with weather forecasts: algorithms are outlined by equations explicitly describing the rules of temperature, pressure and atmospheric speed evolution. These black boxes do not foresee by analogy to the past. Instead, they describe the future in a generative manner. In this chapter, we cover how this paradigm is opening up new tools of predictive powers in biological and social contagion, trends and knowledge spreading. By means of sophisticated models reproducing virtual worlds on a computer, where every individual can be explicitly simulated on the basis of personal attributes and social interactions, we are at the frontier of forecasting's power to predict the spread of diseases like Ebola or Zika. The chapter will also look at how it will be possible to extend these approaches to study different scenarios, from the evacuation of a stadium to the management of a thermonuclear attack.

6. A better world – with data and algorithms

Algorithms and digital data can offer incredible opportunities for real-time analysis and prediction, specifically in the fields of sustainable development and humanitarian aid. Many institutions and research teams are paving the road for a future where big data and algorithms will be deployed in a safe and sensible way – similarly to a public asset: it could go from predicting migration flows to foreseeing the impact of climate change. This chapter will discuss how investigators are exploiting the tools offered by big data science and predictive analysis to improve socially relevant systems, programs and politics.

7. The dark side of the force

Algorithms are fueled by data and pose some ethical issues related to consensus, property and privacy of information. And that's not all. Algorithms and artificial intelligence can generate a tremendous power – the power of prediction – while raising unique ethical challenges. They can be used to understand how to channel flaws in information and change the way information is conveyed according to the individuals we wish to reach. Suddenly every predictive model becomes a marketing strategy, perhaps a political strategy, or even information and disinformation strategies. Which information must be considered proprietary information? And which must be public knowledge? How can democracy or the administration be turned into a society where the one who is algorithmically more powerful is better at forecasting? This last chapter deals with the dangers of a world gaining control and predictive abilities, down to the individual level, and ways to prevent what should be a «force for good» from turning into a dystopian nightmare.

Conclusion: *A short story about a typical day in the future: everything is decided by algorithms that use data we supply through the «smart» objects surrounding us.*

Artificial Seers

(Excerpt from the Introduction)

> «Calchas, the seer, shall rue beginning the sacrifice with his barley-meal and lustral water. Why, what is a seer? A man who with luck tells the truth sometimes, with frequent falsehoods, but when his luck deserts him, collapses then and there.»
>
> EURIPIDES, *Iphigenia in Aulis*

In early 2016, Zika was taking South America down. The World Health Organization spoke of an «emergency of international concern», the highest level of alert. The Summer Olympics to be held in Brazil were just around the corner: 17,000 athletes would compete, while Rio de Janeiro was preparing to welcome over half a million visitors. A contagion that would compromise the future of hundreds of children – inside and outside of the Latin-American borders – was highly feared. We knew that the virus was at the root of what might have seemed like a regular cold, with headaches, rashes and joint pain – symptoms that usually don't raise any concern. But Zika has grave implications when a pregnant woman is affected by it, because the unborn child might suffer neurological issues – including microcephaly, a birth defect in which the head of the baby doesn't grow as much as the rest of the body, stunting brain growth. What we didn't know was where and how the virus would spread during those months, as well as the risk of infection for those who were going to Brazil and then traveling to other parts of the world. My research team at the Northeastern University, in Boston, as part of CIDID (Center for Inference and Dynamics of Infectious Diseases) was given this task. We had tested the system back in 2014, when we were able to outline the evolution of an Ebola outbreak without having to pack our bags and go to the hub of infection. We simply had to get access to supercomputers, feeding into them sophisticated algorithms with a tremendous amount of data, and processing many figures for several months. But this time the challenge was different – even more complex.

One rarely contracts Zika through sexual contact. More often, the infection takes place through what we call a «vector», or disease transmitter: in this case, Aedes mosquitoes. This small insect has a key role. When the insect feeds on the blood of an infected person it might catch the disease and spread it when it bites again. It becomes the link through which the pathogen – the Zika virus – spreads from the sick to the healthy. For this reason, we had to include those mosquitoes in our new models. We had to take into consideration geographical distribution as well as the number of specimens, a task that we thought might have been impossible, since mosquitoes don't fly around with phones in their pockets. However, when we combined climate information with data obtained from specific traps, we created maps of their distribution that were within a few kilometers of accuracy. This information could be integrated with figures on the human population to create models based on the interactions between mosquitoes and humans, describing the spread of the virus among the two populations. When doing these calculations, it is very important to consider our own human mobility, because we humans are the ones that can travel thousands of miles in just a few hours. In other words, we are the ideal way to export the virus to wherever there might be mosquitoes capable of transmitting it. We had plenty of data, since we accessed an aerial and terrestrial mobility database that even monitors displacement linked to commuting or daily errands of people living in over 190 countries in the world.

The first results were a bust. The computer calculated that 80 percent of the population would be infected, an overestimation. Then my intuition kicked in. The formulae I wrote on the window pane in my office are still there: we had to include the socioeconomic condition of the individuals. A person living in an apartment in the city center – or in any house equipped with fly screens – is less likely to get sick than those living in poor suburban areas. We needed more data, more elements to feed the algorithms. This evaluation led us to some realistic estimates, which were then confirmed by what happened in subsequent months. We made a projection of the future evolution of the epidemics in 2017 and 2018 as well as an estimate of the spread of the virus all around the Americas since its introduction in Brazil, which was roughly between the end of 2013 and early 2014. This work helped to support effectiveness tests on possible vaccines.

Epidemiology is just one of the fields that can benefit from this approach, which is somewhere between data science and complex systems. The same methodology is potentially applicable to every other branch of society. It might seem like science fiction, but having a sneak peak of the future no longer belongs only to visionary writers. Psychohistory, a branch of mathematics describing human behavior en masse, once described by Isaac Asimov, has now become a reality. Scientists from all around the world are working on developing models that can predict flight delays, election results, or the odds of suffering from cancer. There is room for improvement, but the techniques are getting increasingly refined. A desire as ancient as mankind is almost within reach: a future without secrets, a future we can control.

At first, the ambition to rule over the future was the driving force behind the science-based approach to nature. Astronomy – the most ancient among natural sciences – derived from the will to make sense of the phenomena that were once considered fearsome: solar and lunar eclipses, for instance. Galileo paved the way by inventing the telescope, an object that allowed him to illustrate planetary movements of the planets with geometrical models. He laid the groundwork for Newton's Law of Universal Gravitation and the Laws of Motion, up to the great success of Maxwell's Equations – the foundation of classical electromagnetism – that at the end of the 19th century made us feel all-knowing. By knowing the origin of the cosmos, its evolution seemed to hold no more secrets. That paradigm was slightly changed by quantum mechanics, which forced us to think about the world in terms of probability. On the other hand, it also opened a window on understanding how the atomic and subatomic worlds work. An intellectual *tour de force* that came culminated on December 2, 1942, when Arthur Compton, one of the great contemporary physicists, made a peculiar announcement: «The Italian navigator has just landed in the new world» he excitedly said. It was a code expression indicating that Enrico Fermi and its colleagues at the University of Chicago succeeded in making Chicago Pile-2 work. It was the first atomic pile in history to produce a self-sustaining, controlled atomic chain reaction.

Contemporary society has been shaped by three hundred years of scientific successes that enable us to predict the weather or the occurrence of eclipses in the coming years. Such knowledge is just a few clicks away. What has been missing is an application of the same kind of scientific approach to daily life: a physics of social systems. Of course, this has occurred to others. The first one to outline a quantitative theory on this was Adolphe Quetelet. He was inspired by the success achieved in classical mechanics and used probability and statistics to characterize society and develop the concept of *everyman*. His approach had two main problems. The first was a theoretical issue. Trapping humankind – man seen as some sort of social atom—into a series of equations seemed counterintuitive: we are too complicated to be trapped by some formulae. The second issue was a practical one: there wasn't enough information to characterize our diverse relationships.

In the last few years, two great revolutionary men have changed this framework. The death of the queen bee myth made society less enigmatic, and downsized the importance of the individual as chief architect of the future. We learned that the complex hierarchical structures we see in anthills do not come from leadership, but from collective phenomena emerging from a great number of specimens performing very simple interactions. The same is true for synchronized flocks of birds and human relationships: the spread of a fad as well as the polarization of opinions can be understood and translated into formulae. Moreover, the «big data era» has given us a laboratory in which it became possible to experiment and collect data about individuals and their interactions. We are hardly aware of the fact that every day we provide a series of personal information. Researchers can use that information to analyze our society. Just to get a sense of this, it is estimated that every day two exabytes of data are produced. That corresponds to a pile of DVDs as high as four Eiffel Towers. If the telescope has allowed us to look closely at the planets, these digital traces laid the scientific groundwork for psychohistory. Our movements can be tracked through our mobile phones; what we search for online outlines our preferences; the sums we pay using our credit card indicate our living standards. Not long ago, this amount of information, and how it captures an almost complete picture of our daily life, was unthinkable.

This information does not represent a new branch of knowledge, though. To misquote Poincaré, a famous physicist and mathematician, a stack of bricks does not make a house, so data accumulation can't be considered science. This is where algorithms come into play: they give data predictive power by canvassing, filtering and making sense of them. Simply put, algorithms can be defined as a series of precise instructions and mathematical equations we use to find associations, identify trends, infer the laws and the dynamics underpinning phenomena such as contagion, the spread of ideas or market trends. Although algorithms have ancient roots, software applications translate then into instructions a computer can follow. They have already replaced us, when it comes to our preferences. After learning our habits, an algorithm suggests products for us on Amazon. On social networks, another algorithm determines what posts we see and analyzes our «likes». Many algorithms are based on automatic learning models – machine learning – and take advantage of statistics:

they search for similarities and repetitions occurring in the data, allowing us to anticipate potential future behaviors. Even our brains unconsciously use statistical algorithms. For instance, we can all forecast that temperatures in August will go from to 20 to 30° C. This comes from a knowledge based on the observation of how seasons occur year after year. To make such a prediction is fairly simple: we don't have to know much about the basic principles governing the seasons or the meteorological system and so on. But when it comes to complex systems, to identify repeated trends, cycles or specific temporal associations is much more difficult. This is where machine learning comes into play and helps us grasp what we can't detect.

Then there are predictive models that not only discover statistical associations, but are actually based on a core of equations describing the evolution of the system through well-defined principles and laws. To again use the same example, we can understand the changing of the seasons, starting from the laws regulating the motions of the planets and the position of the Earth around the Sun. These are the models used in weather forecasting, which – with a few exceptions – we are getting better and better at. In the last few years, this approach has been extended to biological systems, to the spread of epidemics and to other phenomena revolving around the social atom. For several years now, we have been capable of recreating synthetic worlds using a computer. These virtual worlds statistically replicate real ones thanks to databases like LandScan – a project sponsored by NASA that can make estimates of the global population up to an accuracy of one square kilometer. This grid describing people's geographical position can be coupled with an infinite quantity of information such as age, sex, job and geographical displacements. Imagine having a copy of the neighborhood you live in. Whoever lives in this kind of digital village has the same characteristics of your real neighbors: same number of children, same job and same inclination to travel abroad. Think about simulating a few scenarios in order to know in advance what might happen if a member of the community was infected by a virus or if a flood affected your area. It would be exactly like a videogame, except for one non-negligible thing: what happens in the «game» is what – probabilistically speaking – would actually happen in reality.

This capacity can be used for the best. The United Nations created a specific program called «Global Pulse,» which is meant to deploy big data to draw up development policies and help organize humanitarian aid. One of the projects uses figures coming from social media to grasp the perception of emigrants and refugees. Another tracks the consequences of climate change, while another monitors the implementation of HIV prevention programs. There are a lot of ways in which this information can contribute to collective well-being. In 2016, the Baidu's Big Data Lab – a laboratory dedicated to analyzing data owned by the Chinese search engine – announced a newsworthy study. Scientists claimed that they created a system capable of anticipating the formation of major gatherings in some specific areas which could be used to warn us of danger, to prevent displacements of people, disasters and incidents due to unforeseen gatherings of people—but also repress unwanted protests. This borders on dystopia. Although the mechanism wasn't allegedly meant to detect protest areas – because of its short reaction time – who's to say that one day authorities will not deploy it precisely for this purpose?

In a nutshell, these are the advantages and disadvantages of algorithms. The «dark side» of algorithms must also be taken into account. Greek history has taught us that divinatory abilities equal power. In Greek civilization, oracles and seers were considered infallible authorities: they enjoyed fame and riches; they went to war with armies and were consulted before any decision that could impact life at the *polis*. In Homer's *Iliad* the author assigns a very important task to the prophecy of Calchas, «the best of diviners»: to introduce the conflict between Achilles and Agamemnon, the spark of the Trojan War. Whomever can access data and algorithms now has the same gift. They comprise the contemporary crystal ball, one that can change how politics works. Cambridge Analytica uses the Facebook accounts of 50 million of American citizens in order to profile constituents, to understand if they are more likely to sympathize with Democrats or Republicans, and send them targeted political messages. In Italy, complexity, data and algorithms are strictly linked to *Rousseau*, the direct democracy platform used by the Movimento 5 Stelle. In a book called *Tu sei la rete* («You Are the web»), Davide Casaleggio claims that people, like «ant colonies», can be easily conditioned through the spread of simple messages.

The threat of a world where algorithms can not only see the future, but can actually define it, is getting more and more real. It is impossible to turn back the clock, unless you want to throw your smartphone away and live «offline» - which means giving up every kind of technological interaction. Therefore, being more conscious about the algorithm-based technology that we use is the key to ensuring that this predictive power isn't used to manipulate or control us: it is important to know what these «artificial seers» are actually able to do with information, how can they be used for the greater good, and how should they be regulated before it is too late – and the algorithm gets to regulate us.