Wider than the Sky : social physics and big data

Science has provided reliable knowledge, first about the skies and then about the natural world. Today, the avalanche of data on society drives modelers towards this new frontier. Armed with their mathematical tools, they want to develop a quantitative science of society, which would allow to understand and improve it. Is this project realistic ? To what extent are we predictable social atoms? Is the modeling of society useful, and if yes, for whom? Thanks to the rare combination of academic expertise in physics, economics and sociology, the author presents an insider view of the strengths, weaknesses and dangers of the quantification of the social world.

Pablo Jensen is a physicist at Ecole Normale Supérieure and former director of the Complex Systems Institute in Lyon. He founded the French scientific cafés in 1997 and authored "Atoms in My Cappucino ?! Can physics explain everything? "(Seuil, 2001). He has published a hundred articles in international journals in physics, economics and sociology. Presently, he works with social scientists to create formal tools that enrich our understanding of society.

I - Building a common world through science?
Historical introduction and outline of the book

II - How the natural sciences build reliable knowledge
1. Why are scientific explanations robust? Natural sciences allow to control the world by discovering, through manipulations in laboratories, stable relationships.

2. Did Galileo discover the law of falling bodies? His inclined plane represents the first laboratory. It allows to transform free fall motion into a movement that satisfies three essential criteria of modern science: it can be recorded, it is reproducible and results from a single cause.

3. Predicting climate. Weather predictions are often seen as a proof that complex systems can be modeled. We show that this success can be attributed to physical laws of conservation, which are robust in space and time, combined to strong institutions centralizing data and models.

4. Physicists' atoms. In practice, atoms are the entities that keep track of stable relationships within the transformations of matter. But they are not really "atomic" and do not explain the properties of matter.

III – Making virtual societies
Inspired by the success of natural science models, researchers simulate societies in their computers, to understand our own. A detailed study of several examples shows that simple models can be conceptually useful, but more complex ones struggle to become relevant.

1. Simple Models

a. Do segregated neighborhoods imply racist residents ?
b. Opinion or opinionated models ?
c. Economic competition on the beach

2. More realistic models

a. Are we complicated ants?

b. Actions under influence
c. Modeling Ebola epidemics
d. Can models predict economic growth?
e. Does “more data” mean “better predictability” or “more noise”?

3. Why social models are not reliable. Three main factors explain the fragility of social models : The large number of human actions involved, their shakiness, and the reflexivity of social actors.

IV – Analyzing our society
Since virtual societies are not reliable, let's analyze ours, to understand the causes of social phenomena.

1. The birth of statistics: What can we learn from the surprising stability of social averages ?

2. Complex explanations: Are there mathematical tools that can handle the complexity of society ?

3. Is big better? More data do not always increase our understanding

V - Quantifying society
Quantitative indicators try to make social features more objective, to coordinate and legitimize our actions.

1. Physics and the distribution of retail stores

2. A moral thermometer? Comparing the soundness of physics' and economics' indicators (temperature vs GDP)

3. Objectivity and objections. Indicators can be useful even if they are not objective

VI – The constitution of society

1. Are we social atoms ? The analogy is misleading because humans do not have stable internal characteristics which would make their actions predictable

2. The politics of context. Splitting social reality between stable entities and a “context” has political consequences

3. The whole is smaller than the parties We can now understand why society is not « made of » individuals

Conclusion: How can science help building a common world ?

I - Building a common world through science?

Cementing society

History shows that all societies have some common values, some beliefs, placed beyond individuals and cementing the social body. This is also true for modern societies: The US proclaim "In God We Trust", the UK "God Save The Queen" and the French Republic protects "sacred" human rights. In the course of history, order was guaranteed by rites, religions, or the great invention of ancient Rome – an autonomous legal order. Today, science plays a major role in the legitimization of social order. To understand how this has happened, we must go back to the beginnings of modernity.

For centuries, Gods had lent legitimacy to the ruling authorities. Then, in the 16th century, Europe was devastated by religion wars, showing that religion could no longer found political authority. Social order had to rest on a human basis. But which one? Inspired by the new science of motion, the English philosopher Thomas Hobbes wanted to found political philosophy on solid empirical bases. As Galileo had done for the fall of objects, it was necessary to establish by observation the true human nature, in order to deduce from it the appropriate organization of society. In his *Leviathan*, he used a natural experience that he had witnessed in England: the absence of strong government. Without authorities regulating social life, Hobbes assumed that men were in their "state of nature", revealing their deep tendencies. The state of permanent conflict which then prevailed led Hobbes to postulate a that "in the nature of man, we find three principal causes of quarrel. First, Competition; Secondly, Diffidence; Thirdly, Glory. The first maketh men invade for Gain; the second, for Safety; and the third, Reputation". This leads to a state of permanent war and generalized fear. Hobbes inferred that, in order to attain a pacified society, every one had to yield his rights to a dictator, a sovereign power placed above all. This gloomy vision was contested, but the idea of a human nature founding political philosophy survived. As we shall see, economic theory always rests on this idea, well summed up by John Stuart Mill: "The laws of the phenomena of society are, and can be, nothing but the laws of the actions and passions of human beings united together in the social state. Men, however, in a state of society are still men; their actions and passions are obedient to the laws of individual human nature. Men are not, when brought together, converted into another kind of substance".

At a very abstract level, there exist common human tendencies, linked to our biological constitution, such as those detected by Hobbes. But we shall see that these tendencies do not allow legitimizing a specific social organization. Because the relevant human behavior remains plastic, strongly determined by education and socialization: biologically identical humans have built societies as different as ancient Egypt, hunting tribes or modern societies. As Simone de Beauvoir stated: man is a "being whose being is not to be".

Order by numbers

What could provide the basis of social order, if a universal human nature cannot be found? At the beginning of the 19th century, an extraordinary discovery provided an alternative. Suicides or crimes are unpredictable at the individual level, but become remarkably constant when aggregated at the national scale. The Belgian astronomer Adolphe Quételet proclaimed that these "statistical" regularities would found a "social mechanics" capable of governing human societies, and as rigorous as the celestial mechanics of Pierre-Simon de Laplace.

At first, "statistics" (of the Italian stato, state) designated the set of qualitative knowledge useful to govern a country. Then, in the 19th century, European states transformed their territories and populations to make them governable from a center. They counted their populations, measured agricultural production, to improve tax collection or enlist the soldiers. This required the establishment of a legal and physical infrastructure, an investment similar to that of a road or a rail network. European states generalized supervisory tools which now seem self-evident, such as maps, homogeneous measurement units and language, stable family surnames... This control was partly carried out by force, but it was favored by the benefits of modernity such as social inclusion or hygiene. The extension of state power between 1800 and today can be illustrated by a single figure: their budget increased from less than 10% to about 50% of the wealth produced.

The scientific elite invented mathematical tools capable of taking advantage of social data for controlling populations. Laplace, who was Napoleon's minister in 1799, developed different approaches to estimate the French population on the basis of piecemeal data, because it was difficult and costly to carry out an exhaustive census. He assumed that the number of births per inhabitant was almost constant in the country, a hypothesis which he tested in some thirty representative regions. It was then sufficient to count the number of births, which was well known through the parish registers, in order to obtain an estimate of the total population.

The multiplication of social regularities revealed by the avalanche of social data struck the minds. The mathematician Siméon Denis Poisson compared "the almost perfect invariability" of the proportion of convictions in criminal cases to the fixity of Newton's law of inertia. Quételet, a tireless propagandist of “social physics”, helped creating international statistical societies, bringing together professionals and amateurs, doctors, hygienists, actuaries and social reformers. Thomas Buckle, the author in 1857 of a very popular history of England, summed up this enthusiasm: “In a given state of society, a certain number of persons must put an end to their own life. This is the general law […] and the power of the larger law is so irresistible, that neither the love of life nor the fear of another world can avail any thing towards even checking its operation”.

Note that among the readers of Buckle and Quetelet one finds James Clerk Maxwell, one of the most influential physicists of all times, already well known for his work on electromagnetism. In 1859 he published the founding article of statistical physics, showing how to deduce the properties of a gas from those of its supposed constituents, the atoms. He drew inspiration from Quetelet's approach, looking at aggregate quantities, instead of trying to compute individual atomic trajectories. This allowed him to find the distribution of their velocities, and to deduce many gas macroscopic properties, such as pressure.

Planning hope

These statistical regularities gave birth to a new entity, society, endowed with laws that seemed to transcend individuals, profoundly changing both knowledge and political action. A new science, sociology, was created to study it. Through the centralization of data, society seemed to be observable from the outside. This lead the government to conceive itself as a social engineer, manipulating populations as physicists do with natural objects. At the end of the 19th century, states routinely collected data on wages, working times or household budgets, to organize the welfare system. In France, wartime planning lead to a tight collaboration between statisticians, academics and politicians as early as 1914, leading to the systematic economic planning after WWII. In the US, the Great Depression of the 1930s legitimized the creation of federal systems of social regulation. Centralized administrations relied on social sciences experts, leading to the development of modern statistics, i.e. sample surveys, various measures of inequality, econometrics, and the use of computers since the 1940s. The idea of a "representative sample", the basis of current surveys, was developed by George Gallup's firm. In 1936, he contradicted the prediction obtained by a survey of two million readers of the magazine Literary Digest. Thanks to a poll involving twenty times fewer people, but more representative than the readers of the conservative magazine, he managed to predict Franklin Roosevelt's victory. Second World War gave decisive momentum to the welfare state.

States kept on centralizing more and more data, building national databases on incomes, fingerprints or medical expenses via social security records. In practice, society was managed by economic models predicting future growth by combining aggregate household savings, businesses confidence and investment data, or by engineers planning new highways using socio-economic data to compute expected traffics. Centralized planning has led to undeniable progress, such as the generalization of health and education to the whole society. It has also provoked disasters, when the universalist pretensions of planning were combined with an authoritarian central power. This was particularly the case in a number of countries where the desire of socialist-inspired governments to bring the benefits of “modern life” to “backward” peasants, has led to a collectivization that resulted in famines. In *Seeing like a state*, James Scott describes in detail the case of Tanzania, where between 1973 and 1976 five million peasants were forced to settle in large, standardized farms, inspired by the Soviet kolkhozes. This organization, totally unsuited to local agricultural conditions (nature of the soils, family situations...), transformed the experienced peasants into disqualified workers, leading to a catastrophic fall in production. This "high modernism" was a vision shared by elites of all political sides. It declined from the 1970s onwards for multiple reasons, partly related to the very progress it had engendered. The complexity of the economy made planning more difficult, and raising the education level led to an emancipating individualism which challenges centralized plans.

Building things that hold

Science, however, still represents one of the most powerful institutions to legitimate common values. Hence the importance of the attempts of physicists and other researchers who, armed with their mathematical tools, seek to create a mathematical science of society, by using the new avalanche of social data. Judging the value of these efforts is one of the main subjects of this book. For this, we must first understand the origin of sciences' legitimacy. Scientists readily propose a realistic epistemology: sciences are objective because they discover the real world, which is what it is, whatever different individuals or cultures may think. Galileo "discovered" the true law of fall of bodies four hundred years ago, which is still taught in schools today. One can of course object to science "applications", but pure science is neutral, as it merely reveals the pre-existing world. Genes are there, whether we like it or not, but we can challenge GMOs. This is a simple and probably comfortable solution for researchers because it legitimizes their knowledge, while exonerating them from bad "applications" - even if one may note they usually avoid doing so with the beneficial ones. The first chapters of the book analyze natural sciences in action, to challenge this realistic epistemology.

But it is undeniable that sciences construct, according to Alain Desrosières' happy expression, things that "hold", in the triple sense of "who are solid" with respect to the objections of colleagues, who "hold each other" by constructing coherent sets of knowledge, and who "hold men", by leading to agreement around a common world. We will see how climate science, thanks to the titanic work of a large scientific community, has created a virtual Earth, which reproduces its atmosphere in a realistic way, leading to climatic predictions that "hold", in the face of the numerous and powerful opposing interests. This, in principle, enables humans to form a common knowledge to respond to the expected catastrophe.

The recent avalanche of digital data excites many modelers, who want to use it to reproduce society in a computer. For example, the FuturICT project claimed that many of the current problems, “the financial crisis, social and economic instability, wars and epidemics” are linked to human behavior, “yet we know relatively little about how our society works”. By combining complexity theory and social data analysis, FuturICT will allow to “understand and manage our connected world”. The project included a “Living Earth Simulator”, to explore the effects of different policies, linked to a “Planetary Nervous System”, a global sensor network centralizing massive amounts of data on individuals and technological infrastructure in real-time. This project succeeded in slipping among the six finalists to win a European funding of one billion euros. It revived the old idea of Quételet, as one of its objectives is to “reveal the hidden laws and processes underlying societies”, as physicists have done for matter. These laws would provide the ground to simulate the different scenarios, in order to choose the “best one”. This approach has proved fruitful in the natural sciences: Using the laws that govern the atoms, physicists have built virtual crucibles to explore the properties of original materials quickly and inexpensively. They can imagine original alloys and test whether these are able to transform CO_2 into fuels, solving two environmental problems at once. But this approach cannot be extrapolated to society: we are not social atoms!

Is formalization useful?

FuturICT represents an extreme case of a booming approach: the creation of virtual societies to understand and manage ours. We will analyze specific examples, such as economic models of the labor market, or epidemiological simulations assessing the risk of Ebola world pandemics. We will discuss their successes and limitations in detail, and draw some general conclusions.

There are many other ways to build shared knowledge using quantitative tools. Since the 19th century, statistics has developed mathematical techniques capable of analyzing real data, to reveal the causes of social phenomena and attribute responsibilities. Take gender pay inequalities: Do they arise because of direct gender discrimination, or are they merely due to differences in working time or professional qualification? How can researchers state that fine particles “kill 48,000 people in France each year”, when none of these deaths can be directly observed? What is the impact of reducing working time on overall employment? The legitimacy of public action depends on the reliability of these mathematical analyzes.

A third way of building shared knowledge about society is to transform a complex phenomenon into a number. Gross domestic product (GDP), high school rankings or the number of crimes solved by a police department only retain some aspects of reality, to build an “objective” point of view, beyond particular perceptions. The growing use of this quantification method is related to the general political atmosphere, following the general trend revealed by Alain Desrosières. By rating and comparing individuals and organizations, these “benchmarking” methods, imported from the corporate world, are in line with the logics of generalized competition between isolated individuals, driven by incentives and prizes. A detailed comparison between the “moral thermometer” (GDP) and physicists' temperature will show why physical indicators are far more reliable than social ones.

The detailed study of these examples will lead us to the central questions of this book: What is the value of the quantitative approach to social systems? Why is modeling so difficult? After all, one could imagine that, as society is our own creation, these sciences would be easier than those that deal with objects as distant and exotic as galaxies or atoms. And yet, social sciences are the real “hard” sciences! To answer these questions, we must first understand what makes natural sciences robust, going beyond simplistic ideas about the “discovery” of the real world, or a mathematical essence of the world. The first part of the book explains, in a realistic way, the reliability of the quantitative approach of natural sciences.

Further reading

- Hobbes' Leviathan (1651) is available online at https://archive.org/details/hobbessleviathan00hobbuoft. The quotation from JS Mill is taken from his *A System of Logic, Ratiocinative and Inductive* (1843), and that of Simone de Beauvoir from her *The Ethics of ambiguity* (Gallimard, 1947).
- Essential reading on the history of statistics: Alain Desrosières, *The politics of large numbers. A History of statistical reason* (Harvard University Press, 2002)
- How centralized states manage their populations: *Seeing like a state,* James Scott (Yale Univ Press, 1998)
- On the links between statistical physics and social sciences: Th. Porter *A statistical survey of gases: Maxwell's social physics*, HSPS 12, 77 (1981)