

The Networked Society of the Future

(Wolfgang Pietsch, "Electric City" Conference, London, 7.12.2012)

In this presentation, I will be talking about a theme that has repeatedly come up during the conference, namely the impact of big data on the social realm. I will not address the many opportunities that modern information technologies bring to the social world, but rather what I consider the most fundamental challenge. This challenge results from the rise of a networked society, i.e. from a society in which increasingly everything is linked with everything else. In the near future, the internet will expand from a network linking just computers and their users to a network connecting ever more areas of the social and the physical world. In fact, this is already happening. Cell phones, cars, buildings, devices in your home, at work – everything will be equipped with sensors, interfaces, radio transmitters and thus be integrated into the larger internet. This internet of people and things, the integration of the physical and the social world into one single network, permits on an unprecedented scale the collection of ever larger data sets correlating all parameters of human life.

I will argue that these data sets allow, maybe for the first time in history, to establish a substantial body of causal knowledge in the social sciences. The internet and its extension to the internet of things render obsolete the traditional arguments against causal knowledge in social science. The algorithms are much better equipped than pencil and paper to handle the complexity of social phenomena and the internet increasingly allows for systematic, controlled experimentation in the social realm. Owing to the very nature of causality, this body of knowledge which is being amassed by machines in the machines will allow for an increasing predictability of and control over the social sphere.

Let me start my argument with the analogy between society and another complex and chaotic system, the weather. What you see in the graph on the left is a depiction of the considerable improvement of weather forecasts over the last decades. The different lines denote forecasts for three, five, seven and ten days, both for the northern and the southern hemisphere. The slope denotes the improvement in weather predictions. A slow and gradual change has occurred in recent decades, which nevertheless means that for example today 5-day predictions are as good as 3-day predictions were ten years ago. Essentially, there are three different factors that have led to this increase in accuracy: more weather data, increased computing power, and better models. Since the change is happening so slowly, for many of us this quite spectacular improvement has gone largely unnoticed.

Remarkably, we are facing today a similar situation with respect to social phenomena, where the same three factors are present. Over the past decades, enormous data sets have been collected mostly by large internet companies like Google, Yahoo, or Facebook. This data is analyzed by ever more powerful computers employing automated modeling techniques. The data sets allow for an improved ability to make predictions about social phenomena – predictions which are getting better just as imperceptibly as weather forecasts are. Google, for example, is engaged in such 'social forecasting', trying to predict the movement and

needs of internet users. On Google's Trends platform, the search history of specific terms can be looked up like in this case "internet of things" and how this term correlates with geography or news items. Google's algorithms also hypothesize about the popularity of search entries in the future, the dotted extension of the line, aiming at predictions in the as yet limited social world of the internet.

There is a simple formula for turning the ability to predict into an ability to control. This formula has been known at least since the times of Francis Bacon, the intellectual father of the Royal Society and long considered the founding figure of modern scientific method. Bacon famously wrote that "human knowledge and human power meet in one". Less known is that he explicitly justifies the link by referring to causal knowledge, "for where the cause is not known the effect cannot be produced". Thus, if the causes of a phenomenon are known, it becomes predictable. If in addition one can manipulate these causes, then one also has control over the phenomenon. Plausibly, the internet and the internet of things allow for just that, first, for the collection of large correlated data sets which are crucial for a causal analysis of complex phenomena and second, for an increasing ability to intervene on the relevant causal parameters from the distance and on a large scale. This new kind of control over the social sphere, if used systematically, might at least in some contexts drastically reduce our freedom to choose and thereby have the stupefying effects that Richard Sennett has been talking about.

Let me give you an example that such mechanisms have successfully been put to work, that large correlated data sets indeed allow for better prediction and control in the social sphere. In his recent book 'The Victory Lab', journalist Sasha Issenberg reports how data-intensive methods have in the last decade changed the way how American election campaigns are managed. In a review of this book, James Surowiecki, a columnist for the New Yorker Magazine, describes how the Obama campaign in 2008 employed "a custom-built, constantly evolving algorithm that incorporates hundreds of variables in order to predict any given voter's allegiance and level of enthusiasm." He continues that "to campaign operatives it sometimes seemed that the algorithm knew what voters thought before the voters themselves did." This is a long and complex story, but there is good reason to believe that on the basis of large data sets it was possible to identify some causally relevant parameters which subsequently could be manipulated. Apparently, this allowed for a small but definite degree of control to make people vote or even influence for whom.

Remarkably, the issues that arise in this context were already fiercely debated in the 19th century involving the most prominent social scientists of the time. In a way, they were facing a situation that is comparable with today's developments, namely an explosion of social data if still collected by hand and not yet by machines, and the parallel invention of a number of powerful statistical tools. At the same time, there was a virulent debate to what extent human actions are determined by social boundary conditions, if the social sphere allows for causal laws just as the natural sciences do, and conversely, how much man is free in his choices. Is it true for example, as Adolphe Quetelet, the Belgian polymath and one of the

most influential figures in this debate, claimed that “society prepares the crime and that a guilty man’s crime is just the fruit of the circumstances in which he finds himself”. In the 19th century, the debate on how much human behavior is determined by causal laws ended without result, mainly because all candidates for such laws turned out much less stable and reliable than originally expected. Back then, there just was not enough data. But possibly, this has changed with the gigantic data sets that are now being automatically collected and analyzed. Even though it remains uncertain if this data suffices to match the complexity of most social phenomena, the same issues that were discussed in the 19th century need to be addressed again. To what extent can the enormous data sets of the internet be used for a control of the social sphere? This is not so much a question if man is free or not, but rather at what instances we are free and to which extent. Free, i.e. determined not by the external conditions, but by our own internal conscious deliberations.

Let me thus conclude: As imperceptibly as weather forecasts have improved in recent decades, a new kind of social science has arisen that can only exist in large data centers as the one of Google which you see on the right. You may call it the ‘social science in silico’, in the computers. It is made possible by the large data sets of the networked society which increasingly link every aspect of human life. This novel kind of science comprises an almost infinite number of dynamically evolving correlations and causal laws. Like other sciences, it allows for predictions, for a certain degree of control, and it is to some extent explanatory. But in other respects, this science is unlike any other science that is currently taught in schools or at universities. It is a science that has to stay in its digital cage as it does not fit into the limited and static environment of conventional textbooks or even our human brains.

Already today, but increasingly so in the future, this science will allow for the control of certain parts of the social world. It is no exaggeration to state that slowly a new instrument of power has arisen. And this situation requires a debate on how this power should be used and who should be allowed to use it. To make it clear, this is not about privacy, it is not about personalized data. Rather, it is often irrelevant if the data is linked to a specific name or not. We need a debate how this social science in silico can be employed for the human good. The status quo is at least problematic that access is restricted to a small number of high-tech companies who largely use it for their own economic success.