



The Green Templeton Lectures 2016

LIVING BY NUMBERS: BIG DATA & SOCIETY

BRIEFING NOTE



INTRODUCTION

What is the digital future, and how will 'big data' significantly change our lives and how we conduct research?

In the 2016 Green Templeton Lectures, four high-profile speakers – Susan Halford, Helen Margetts, Tim Lang, and John Bell – addressed various aspects of this timely question.

Drawing upon their own projects and experiences, the speakers discussed in turn the impact of big data (the availability and analysis of large datasets) on social research, collective action, food, and medicine.

Such datasets have arisen from the increasing digitisation of everyday life. Due to their size and complexity, they have raised numerous challenges, as well as opportunities and potential benefits relating to data collection, communication and usage, and methods of analysis.

Twitter & Social Life: Tales from the Frontline of Social Media Research

25 January 2016

Professor Susan Halford, Director of the Web Science Institute at the University of Southampton, opened the lecture series with an exploration of the challenges and opportunities afforded by social media data for social scientists.

A profound shift has occurred. Digital technologies are no longer just a research tool; the data they generate have become the substance of research. Social scientists do not have to rely upon interviews and surveys to understand human behaviour – they now have a ‘telescope’ through which they can directly observe what people say and do in day-to-day life, on an unprecedented scale, in real time and over time. In the words of Bruno Latour: “It is as if the inner workings of private worlds have been prised open.”



“With its 230 million users and 500 million tweets per day, Twitter is the best example of such a ‘telescope’. It has come to dominate social media research.”

However, it has also led to significant challenges relating to disciplinary focus and boundaries, the nature of big data, and associated ethical issues. Twitter is used differently in each discipline and to date little interdisciplinary work exists.

Thus, methodological advances are inhibited, with social scientists conducting small-scale content analysis or large-scale random sampling, underutilising the dynamic nature of the data. They also tend to impose top-down categories. On the other hand, computer scientists capture the dynamics of the data through mathematical networks but do not ground their conclusions in theory.

The second challenge identified by Professor Halford concerns the data themselves.

“Contrary to what the Research Councils UK claim, tweets are anything but ‘naturally occurring’.”

It is possible that up to a quarter of all tweets are produced not by humans but by automated accounts. We must therefore take time to understand the structures and practices underlying Twitter data. Users are constrained by platform limits and their tweets are not simple reflections of pre-existing categories. Tweets are a commercial asset. Twitter allows researchers access to data on their own terms: those who pay more obtain more tweets.

Professor Halford’s third challenge concerned ethical issues. Tweets are created for public consumption, with no thought by their authors that they might be used by researchers. Obtaining informed consent from millions of people is clearly impossible. To meet these challenges, it was suggested that we must develop hybrid methodologies from multiple disciplines, examine how platforms have evolved over time, describe the technical infrastructures underlying the production of data, conduct experiments that compare samples from different levels of data access, and understand how web robots contribute to social networks. New forms of ethical practice must also be developed that more actively involve the users who produce the data.

Big data will undoubtedly change the future of research in every discipline.

Professor Halford felt privileged to be involved in this exciting field, but stressed the need for researchers to exercise caution in presenting conclusions from their analyses of social media posts.

On a higher level, digital data are rapidly becoming a new infrastructure in the modern world: like roads, they will take us some places and not others, and they are the outcome of fierce competition between vested interests.

Politics by Numbers: How Social Media Shape Collective Action

1 February 2016

Delivering the second lecture in the series, **Helen Margetts, Director of the Oxford Internet Institute and Professor of Society and the Internet at the University of Oxford**, focused on the political dimension of social media data. She drew upon material from her recent book, *Political Turbulence: How Social Media Shape Collective Action* (Princeton University Press, co-authored with Peter John, Scott Hale, and Taha Yasseri).



What is the relationship between digital technologies and collective action? What does politics look like in our changing digital world, and how can new forms of data help us to find out?

“Social media facilitate micro-donations of time and effort to socio-political causes. Such tiny acts (eg watching a video clip, sharing an article) are a qualitatively new form of political participation.”

They can scale up to the mobilisation of millions and achieve policy change (the petition on road pricing being the first instance of this in the UK). They have also changed the research landscape, as they represent real-time transactional data that can be modelled mathematically. Political scientists can now run experiments to test the effects of platform design and predict political activity.

Most mobilisations, however, fail to bring about a political effect. Over 100,000 signatures are needed to instigate a parliamentary debate in the UK: 99.9% of petitions do not reach this number. Success, if it comes at all, comes quickly: there needs to be an exponential rate of adoption. The traditional, well-understood S-shaped distribution does not apply to online processes. Thus, collective action on social media is extremely difficult to predict.

Moreover, traditional predictors of political activity (eg demographics) do not work. Personality seems to be more important. Social media can influence individuals’ decisions to participate in political acts, both by providing real-time information on what others are doing and by making their actions visible to others. Different personalities seem to respond differently to social influence: extroverts are more likely to initiate political acts even when there is no social information, while visibility has a greater effect on individualists, who are more easily shamed than cooperatives.

“The dynamic and unpredictable world of online collective action is one of political turbulence.”

Collective action without leaders and institutional support is possible but without any guarantee of effecting change. This partly explains why so many of the mobilisations since the Arab Spring have been disappointing. What are the implications of political turbulence for democracy? Professor Margetts and her co-authors suggested that the phrase ‘chaotic pluralism’ represents what is happening: we have a pattern of competing interests, with distributed influence, but it is far more disorganised, fast-moving, and individualised than earlier pluralism models.

For researchers, much remains to be done. On a methodological level, data science approaches are required to fully understand this turbulence, including mathematical models, experimentation, and machine learning. Professor Margetts suggested that we have to work with the internet companies that produce the data: there are huge sources that we have no access to. We must link big data to individual-level data. And we have to bring multidisciplinary into the mainstream.

Finally, terms like ‘slacktivism’ are not useful. Many more people than ever before are participating in political acts, and such acts do lead to more substantive acts. We should focus on the exciting ways through which we can harness this willingness to participate in our political system. The numbers *do* matter.

Big Data, Food Consumption, and Food Policy

8 February 2016

Moving the focus of the lecture series away from social media, **Tim Lang, Professor of Food Policy at City University London**, explored the significance of big data in the world of food.

In the 21st century, the challenge has shifted from how to produce plenty of affordable food to the transfer of food power among consumers, the government, and corporations. This power concerns information and minds, not just nutrients, bodies, and ecosystems.

What exactly *is* big data in food? There is no clear definition according to Professor Lang, but it is worth noting that academics and corporates have polarised perspectives.



Academics are critical of big data, claiming that it can be used for surveillance purposes and that it creates a power divide because commercial organisations have control. Big data is neither all-encompassing nor reliable; it can obscure and marginalise consumer categories.

Corporate views are mostly optimistic – big data will produce profits, and is unprecedented because of the five Vs: volume (quantity), variety (type), velocity (process speed), veracity (quality), and (monetary) value. The main concern is over the technology that will work best to harness its potential.

“In the world of food, big data is not intrinsically new. Detailed information has long been collected about consumer purchases. What is new is that the language is neoliberal: the market, not the state, is in control.”

Big data in food is an asset to be stripped, and it threatens to reconfigure how we conceive ourselves. Professor Lang opined that we have been reduced to atoms.

Accordingly, much policy discussion about big data in food has focused on the technical power of the system. There is a tension between top-down control and democratic control. The former predominates, whether the goal is to gather mobile data to maximise crop harvests in Africa or to predict consumer behaviour (Tesco’s Dunhumby is the most sophisticated data system for doing so).

This raises significant issues of private vs public knowledge. To echo the Romans: who controls the controllers? Who watches the watchmen? Despite the rhetoric of consumer sovereignty, public policy is not promoting the democratisation of big food data opportunities. A new food citizenship is elusive.

Professor Lang drew attention to the many questions that remain unanswered.

“Can big food data *be* democratic? Who rightfully owns the data? Does it represent a new phase in the battle over the consumer? Is its power overstated? What is the role of big data in and down the food supply chain? What is the role of the state? And how can the workers (consumers) be compensated for their free labour?”

The future of big data in food, however, is promising in Professor Lang’s view. Given the political dynamics of food, this is rich territory for social science research. There are open food data initiatives and emerging policy frameworks; one example is Global Open Data for Agriculture and Nutrition (GODAN), a consortium launched in 2013 to empower farmers in developing countries.

Big data in food also represents a test case for consumer sovereignty. We need laws and regulations that affirm the right to consumer knowledge, in the same way that we have access to our NHS records.

It is important to go beyond the hype to ensure that big food data does not become ‘Big Food Brother’.

Big Data and Biomedical Research: Developments and Implications

22 February 2016

Delivering the last lecture in the series, **John Bell, Regius Professor of Medicine at the University of Oxford**, reflected on the impact of big data on biomedical research.



Although big data have heralded a transformative change in the way we conduct research and understand disease, medicine has been among the slowest of disciplines to recognise the value of large datasets.

There are four main problems: disease is poorly defined, delaying the application of interventions; healthcare systems are financially stressed by demography; innovation is not always applied efficiently; and healthcare systems are often blind to insights about patients that could be revealed by digitisation.

Professor Bell pointed out that our taxonomy of disease has not progressed very far since the 16th century. The organising principle is not what is causing the disease; superficial categories are being imposed: eg heart, liver, and ovarian failure are lumped together, as are irritable bowel syndrome and fibromyalgia. Because our taxonomy does not recognise how highly heterogeneous diseases are, therapies only work for relatively small percentages of people and medicines are not very precisely targeted.

“Big data techniques could result in better understanding of diseases and improved targeting of treatments, as they dramatically reduce confidence intervals.”

Such data would include electronic health records, biobanks, and information generated from wearable devices. Combining them allows researchers to develop algorithms for diagnosis, therapeutic responses, and clinical decision support. A big challenge and opportunity is how to harness longitudinal data on paper (patient records, hospital admissions) and add it to our digital records.

Professor Bell cited some exciting progress in this direction. Genomics England is sequencing 100,000 genomes from 70,000 National Health Service patients; the data platform will not only enable more effective treatments for rare diseases, but should also trigger economic growth.

Digital pathology has enhanced the image analysis of cancer cells through algorithmic models that predict survival. Significant advances have been made in the characterisation of pathogens for diseases like malaria, through clinical epidemiology and large-scale genetic analysis. Research on biomarkers has redefined asthma by identifying inflammatory and non-inflammatory forms, unleashing a tidal wave of new, highly effective drugs.

Such are the powers of precision medicine, which is grounded in the ability to accurately define patient populations.

“The United Kingdom has pioneered much of this progress; our big data assets in medicine are unparalleled.”

We have a number of powerful funding agencies and research institutions: The Wellcome Trust, Medical Research Council, NHS, Higher Education Funding Council for England, National Institute for Health Research, Genomics England, Innovate UK, UK Biobank, and The Farr Institute of Health Informatics Research. According to Professor Bell, England is uniquely positioned to make even more significant discoveries through big medical data.

Oxford University is, and will continue to be, a key centre for this research. The new Big Data Institute at the Old Road Campus in Headington will employ 350 data scientists. They will be from different disciplines: epidemiology, statistics, computer science, and biomedical engineering. This balance of domain and analytic expertise is the way forward for healthcare research; it allows us to generate knowledge from big datasets that will transform our understanding of disease.

CONCLUSION

This lecture series has illuminated the challenges and opportunities presented by big data in a range of domains. The opportunities are undeniable – large datasets have allowed for the discovery of unprecedented insights in everything from political mobilisation on Twitter to drug-resistant tuberculosis.

The challenges are multifold and complex. One key difficulty is access: most big data are controlled by commercial organisations. Should academics engage more with these organisations? To what extent are we able to do so, and what forms can our partnerships take? Another challenge is discipline: to analyse big data, our methodologies need to be much more plural. How can we transcend territoriality and disciplinary boundaries and not only recognise the value of interdisciplinary research, but also actively create projects and academic positions that require it?

Finally, we must critically interrogate the *construction* and *control* of big data information alongside the data themselves. We may be living by numbers, but we still need to understand and develop the rights and responsibilities of the people behind them.

Podcasts of all four lectures are available on the Green Templeton College website at:

www.gtc.ox.ac.uk/podcasts

The Green Templeton Lectures

In an annual series of four lectures in Hilary Term, the Green Templeton Lectures explore a given contemporary theme through a number of different perspectives, whether it be historical, political, educational or philosophical. Designed to be of interest to the College community and the wider world, the series focuses on an aspect of the College's work in human welfare.

High-profile guest speakers have included Dr Navi Pillay, UN High Commissioner for Children's Rights 2008-2014, Dame Helena Kennedy, leading barrister and expert in human rights law and civil liberties, Adam Mars-Jones, novelist, and Sir Michael Rawlins, then Chairman of the National Institute of Health and Clinical Excellence (NICE).

Previous themes have been:

2015 *The World's Child*

2014 *The Tyranny of the Normal*

2013 *Feeding a Better Future*

2012 *States in Crisis*

2011 *Living with the Coalition*

2010 *Uncertainties and Insecurities*

2009 *Addicted to Big Pharma? Reconciling Business, Medical and Ethical Needs*

For more information visit www.gtc.ox.ac.uk/gtlectures

Green Templeton College | 43 Woodstock Road, Oxford OX2 6HG | +44 (0) 1865 274770

www.gtc.ox.ac.uk

© Green Templeton College 2016