



City Research Online

City, University of London Institutional Repository

Citation: Chan, D. (2011). The world is changing – our mindsets must change with it. Big Data Insight Group: 1st Industry Trends Report, p. 12.

This is the unspecified version of the paper.

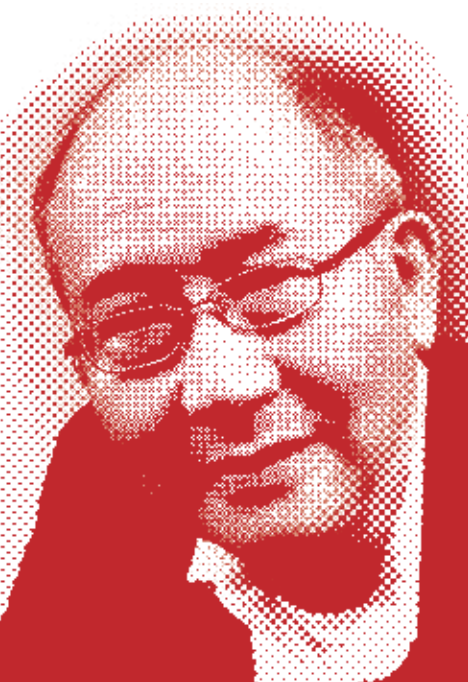
This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/1284/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



The rise in the amount of data we are generating and storing has been meteoric of late. But while the world of data is developing rapidly, the ways that we manage it are not keeping pace. **David Chan** says our paradigms need to change.

A column by David Chan, director, Centre for Information Leadership, City University London david.chan.1@city.ac.uk

THE WORLD IS CHANGING

– our mindsets must change with it

Big data has arrived; there can be little doubt about that. Its arrival has come not only with a meteoric rise in data but with the advances in the tools and technologies we use to analyse it. However, technology is never a silver bullet and before we can reap the benefits of big data we must change our mindset for how we approach it.

Paradigms are the way we make sense of the world. They are the perspectives that we use to understand what is happening around us. But, of course, if we don't have the right information or enough of it, our understandings can be built on falsehoods. For instance, the ancient Greeks understood thunder and lightning to be the god Zeus throwing his thunderbolts. They didn't know about the interaction of warm air, water vapour and static electricity.

The prevailing paradigm for data analytics is static modelling – the search for rules that explain variation in data. We examine randomised data sets and look for patterns between the key points of difference. If we can find these same patterns in other data sets then we know we have a model. We then use that model to make predictions. A 'real world' example of this can be seen in the way that banks weigh up the risk of giving somebody a loan. They analyse the applicant's financial history and lifestyle and try to see whether the patterns fit somebody that usually meets their repayments on time or somebody that defaults.

Static modelling like this is fine, in theory. Indeed, it works perfectly well when there are fixed relationships that are discoverable through detailed analysis. But what happens if what we are seeking to model does not work in the same way?

The 'butterfly effect' is common in mathematics. This is where small changes in the original conditions have hugely disproportionate effects on the outcome. So if you went back in time a thousand years and changed something very small, the world could potentially be a very different place today. In data analysis, the effect can often be that the model we use to gain reliable information out of our data no longer works when applied to the rapidly evolving environment around us.

Unfortunately, these small changes are becoming increasingly common as our data sets get bigger and bigger. Furthermore the sources that we now capture all this data from consist of complex, mutually interacting systems and this makes them prone to feedback. We think of the problem as 'noise' – if you're listening to a conversation and then somebody else talks over them, it's easy to mistake what the person you're speaking to has said.

This 'noise' is especially common now that we are using unstructured data. For instance, if we look at surveys or polls, to what extent are the answers affected by the way the respondents wish to appear?

Although we can account for identified factors, how do we know there are not other factors that have significant impacts? The effects we are trying to predict may well be swamped by other signals.

Therefore we need to change the way that we manage our data and the paradigms that we use to understand it. New technologies like Complex Adaptive Systems are specifically designed to help with noisy data. But without the right approach, they aren't enough.

We need to use big data analytics as a more accurate means of adapting a system to cope with the data being received from these constantly evolving markets. It's new, and not generally well known, but it is the answer. The data for extremely in-depth insight has arrived, and so have the tools; we now just need to change our mindsets to unlock the value they can offer.

Technology is never a silver bullet and before we can reap the benefits of big data we must change our mindset