Big Data, not magic data

Mark Earls and Alex Bentley

Admap

September 2013

warc

Title: Big Data, not magic data
Author(s): Mark Earls and Alex Bentley
Source: Admap
Issue: September 2013

Big Data, not Magic Data

Mark Earls
HERD

Alex Bentley
University of Bristol

Big is not always better. Simplification is often better than precision. Ask appropriate questions of data sets and spot the patterns that lead to genuine insight.

Big Data is, without doubt, the marketing theme *du jour*. It's hard to go to a conference, read a marketing magazine or even sit in a meeting quietly minding your own business without some enthusiastic mention of Big Data. However, being popular is by no means the same as being right or better. Indeed, the marketing buzz around Big Data brings to mind Arthur C Clarke's famous observation that 'any sufficiently advanced technology is indistinguishable from magic'.

To see what we mean, try substituting the word 'Magic' for 'Big': if we only master Magic Data, it will make us all-powerful; the sword of Magic Data will banish all evils (like advertising ineffectiveness and what direct marketers with no hint of irony call 'wastage'); Magic Data will wash away all uncertainties (like knowing what customers really want and what to say or serve at any one time) and remove any need to put that unreliable executive/human judgement of ours to work and on the line. 'Boy, we need to get us some of them Big Data Magic Beans, don't we?'

Let's be clear: we are big fans of data. We're not like those whose response to the 'data firehose' is to retreat inside to instinct and intuition, guesswork, tea leaves and tarot cards. Quite the opposite: we champion evidence-led thinking at the expense of guesswork and speculation; we infuse our work with any data that's available (and spend time finding data that others don't have access to); we 'sweat' it, too, in all kinds of ways.

But in our view, the collective rush to Big Data makes a number of errors – both statistical and practical. And the advocates –

be they the big software vendors, big consultancies, digital media owners or futurologists – are exploiting our tendency to 'magical thinking'. We hope that by shining some daylight on the three myths of Big Data, you might begin to see beyond the sparkle and the whizz-bangs and begin to take a more balanced position.

More is better

Stands to reason, doesn't it? Some data is better than no data, loads of data better than merely some data. Really 'Big' Data must be best of all. But is this really the case? *The Black Swan* author, Nassim Taleb, points out that more data often means more noise and less signal. Pollster Nate Silver writes the same about his predictions for the 2008 and 2012 US elections.

The lure here was, and remains, making sense of the large amount of data that's available. One way to do this is to make predictions at the same large scale as the data. The developments in e-commerce, such as recommendation systems in retail applications, are surely only the tip of the iceberg in terms of Big Data studies aimed at forecasting future behaviour, such as financial and commercial activity, economic trends, epidemics and even crime.

A number of European scientists have made remarkable use of freely available data from Google, etc. To correlate mass behaviour with all sorts of geo-located measures from their devices, such as ambient temperature, noise level, luminance information and energy consumption.

But all of this is at the level of *mass*, not individual, behaviour. Like a telescope surveying billions of galaxies, it works as long as we keep our predictions to the same big scale, such as the way Google search volume predicts financial trends, or new-release sales of films, games, and songs at a population level, rather than in terms of individual behaviour.

More data = better prediction

This big scale often means the prediction can be mundane, as when an analysis of millions of tweets 'suggests that people awaken later on weekends'. If you already noticed that people awaken later on weekends, then give yourself credit as a human tweet-o-meter. This is not a trivial notion; you have observed human behaviour and listened to human remarks virtually every day of your life. That is a lot of observational data, processed by a very powerful computer (your brain) designed through millions of years of evolution to make sense of other people's (or hominin's) behaviour. Your life's worth of observations may be subjectively biased, but they are still 'Bayesian' in the sense of a probabilistic model based on a continually updated model of observed human behaviour.

While Big-Data analytics can predict urban umbrella sales after it starts to rain, or the opening box office of a new movie that everyone is already Googling, human brains can do far more – the standard rom-com is funny precisely because the audience's knowledge of sexual psychology is so sophisticated and nuanced. Indeed, Jonah Berger of the Wharton School and Gael Le Mens of Stanford have argued that subconscious human knowledge of trends can be pervasive and long-lived. Have you ever found your novel idea rapidly ascending the 'most popular' list? That's what Berger and Le Mens are talking about. This is why the next generation of Big Data science is enabling search engines to outsource the most complicated tasks (like face recognition) back to humans.

Big Data science works best when it is not magic, but magnifies our common sense with bigger and more accurate ears. In other words, Big Data science excels at predicting near-future behaviour of a population based on recent past behaviour. One could argue, however, that Nate Silver updated his sophisticated data-rich mock elections right up until the eve of the real election. His virtuosity lay in understanding the data and the complicated electoral process. The larger the sample we have of past decisions, and the shorter time ahead we extrapolate those, the better the predictions will be. Copying recent success is one of the oldest strategies for success that we have as humans. Patterns from Google searches even show that internet users from higher GDP nations search for more information about the future than about the past. Another very productive trail of research looks to predict crashes and collapses using Big Data, revealing a robust set of criteria that demonstrate whether a tipping point is imminent.

But this is still a far cry from prediction of how specific individuals will behave with any certainty. The studies discussed above do not need sophisticated behavioural models, they can just exploit the fact that people generally rely on past observations to make their decisions. If only, we say, we had more of the right kind of data, we'd be able to be much more certain about what individual consumers are going to do next and we'd be better placed to serve them the products and messages that were entirely appropriate to them. Discussions of Big Data seem to always end up in a *Minority Report* future for advertising – direct marketing on silicon steroids.

But is this credible, either? Well, not exactly. While data-mining has helped retailers like Tesco develop the ability to test and learn at scale over time (thanks to the huge transaction datasets that they collect about their customers' purchases with them), even the best of these operations still gets it wrong a lot at the individual level, as we all know from the kinds of wacky 'personal recommendations' we may receive from Amazon. Again, the way forward appears to be to outsource the complicated part to human beings, as Facebook does when it exploits information within the user's social network.

As it happens, much of what we call mass behaviour is not best understood by aggregating up what we know about individuals (no matter how many) but by the interaction of those individuals with each other. This makes it a complex rather than a complicated phenomenon and, as such, strictly unpredictable (as Taleb points out in *The Black Swan*, the bell-curve is subject to extreme outliers that can change the game completely and rapidly). Looking at an individual's behaviour is missing the point here: no matter how many data points you have.

Moreover, as we've repeatedly pointed out in *Admap* and elsewhere, however appealing this Utopian vision of micro-targeted advertising, the truth is a large part of the role of advertising and marketing communication has always been to influence the space between individuals, not just the space between their individual pairs of ears.

Real time = better time

In other words, the more granulated the time series is, the better; the more time points we have in our data set, the more valuable that data is. Well, for some things, yes; but for many things, no. This is partly a question of appropriate scale: for businesses that depend on precision in timing (airlines?) This might be important; but knowing at what minute past the hour Dove shampoo is bought is less useful and less relevant. If, however, it turned out that Dove shampoo was bought more on Fridays and Saturdays than Mondays or Tuesdays, that might tell you something important about when to flight your promotional activity. Different scales of granulation for different courses if you like.

Indeed, appropriate scale is a consideration for data users: more detail is not necessarily better detail. The late Andrew Ehrenberg is perhaps most famous among today's marketers for his work on the 'Dirichlet' – the underlying statistical relationship between share and size and loyalty in many Fmcg markets – but in other circles, he is better known for his work on presenting data. Particularly influential was his insistence that marketers presenting, for example, brand share data to two decimal points serves merely to create a false and distracting sense of variation from the bigger and more important patterns that rounded, cardinal, numbers give us.

This also touches on a bigger issue: a mismatch between too much detailed data and the much larger scale of the objective. This is why we love simple, user-friendly metrics like Millward Brown's AI or Bain's NPS. While some bemoan how infographics reduce the detail that researchers prize, we think this is precisely why such infographics are useful and popular. Ask yourself, are you being too precise? Are you disguising what matters for the scale of your business by showing off how closely you can measure or analyse things?

Implication for marketers

Excitement about Big Data is entirely misplaced: to make lasting improvement to marketing's performance, we have to make better use of the data available to us, and there will only ever be more and more data available to marketers in the future, not less (whatever happens with data privacy).

Here are three very simple things that you can do today:

1. Use the data you have better through pattern-spotting

Pattern-spotting helps make sense of large data sets you're already sitting on (but don't do much with). In our book *I'll Have What She's Having*, we describe some of the patterns that can be easily traced in, for example, sales data, the long-tail distribution of popularity in a market shaped by social learning (copying), rather than independent choice. You may soon be led to very different product research and communication strategies (Figure 1).

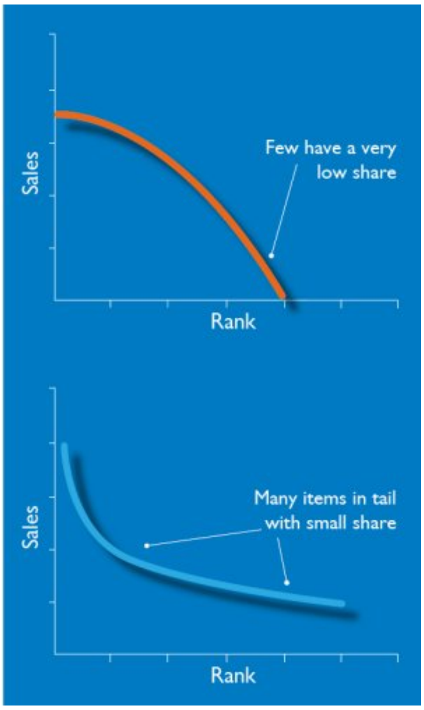

Figure 1: Short vs long tail distributions

Keep it simple: you don't need the kind of sophisticated dashboard described, for example, by Glenn Granger in his recent book *Rain Dancing*; just the willingness to do some simple analysis.

2. Ask yourself more 'what kind of thing?' questions

The value of data lies more in the questions you ask of it than somehow in the data itself; more in what you do with it than what it says on its own. So if you want to make sense of a large amount of data quickly and usefully, chunk it out: ask different kinds of questions of it.

Ask 'what kind of thing is this?' (see 1. Above) straight away so you can link your analysis of the data itself to other learning and experience the organisation has gained – 'What have we learned about this kind of scenario?' 'What are the best and most appropriate kinds of response to it?' use this to generate more useful conversations and decision-making inside the organisation – discussions at an appropriate scale.

For example, we have developed a very simple map of human choice styles based on the kind of patterns discussed above, which help us shape our clients' decision-making about strategy, messaging and communications planning, with clients as varied as The Gates Foundation, Unilever and the Sony Corporation. Having this kind of map – rooted in large data sets – helps land the big insights from the data in these conversations in a way that marketers can act on (Figure 2).

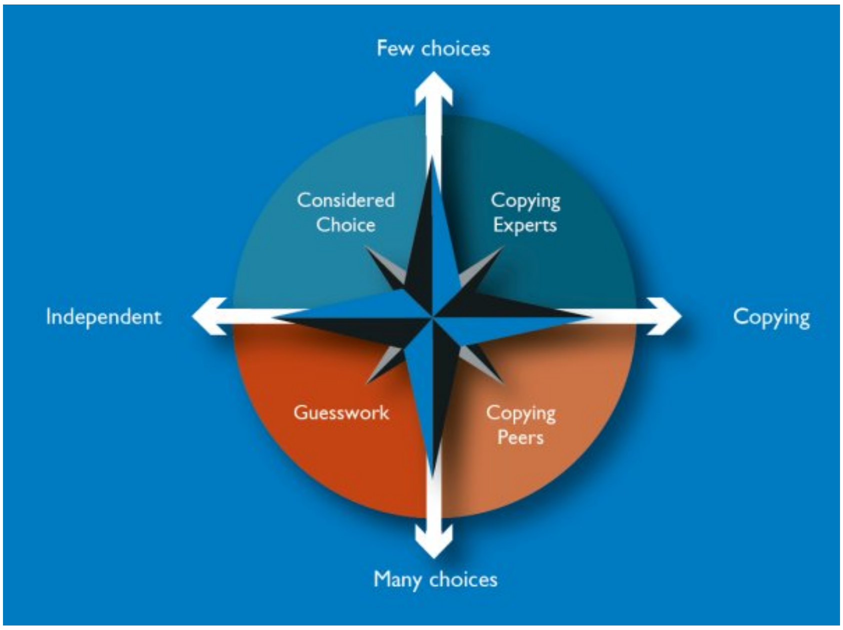

Figure 2: Map of human choice

Adapted from Bentley Earls O'Brien 'I'll Have What She's Having'. *MIT Press 2011*

Our experience (in both marketing and academic circles) is that everyone wants to add nuance and detail to this map, but this is only possible because the map is so simple, clear and robust: it explains much of what is to be explained with ease and leaves only a small amount to be described and debated.

3. Get good at using the free data that's available

Pattern-spotting in freely-available data is often easy and quick. Google provides many of these tools, like Google trends and Google Correlate.

For example, a colleague was recently pitching for the redesign and rebranding of a VIP lounge. Pointing him to Google Ngram (which records the frequency of words in a sample of all the books Google has digitalised) revealed that the word 'VIP' peaked in popularity some time ago and fundamentally shifted the team's thinking about the task in hand.

Knowing from the patterns in search data that a particular idea is spreading through a population because of something to do with the idea itself (rather than just fashion) can guide your strategic thinking more usefully (and cheaply) than any large scale polling exercise. Data become insight into the right sorts of questions.

The bottom line is this: Big Data aren't magic and more are not more valuable. Don't fall for the 'Magic' sales pitch – them there's not magic beans.

About the Authors

Mark Earls is a recovering account planner, whose consultancy and writing (*I'll Have What She's Having*) are rooted in the insights provided by contemporary behavioural and cognitive science.
theherdmeister@gmail.com

Alex Bentley is Professor of archaeology and anthropology at the University of Bristol and deputy director of the Tipping Points Grant at Durham University.