

By Ben Marshall

Putting the Behaviour into Behavioural Analytics

BEHAVIOURAL ANALYTICS IN THE ABSENCE OF THEORY CONCERNS BEN MARSHALL, PARTICULARLY WHEN APPLIED IN THE CONTEXT OF LAW ENFORCEMENT AND SECURITY. HERE HE EXPLAINS WHY.

The term ‘behavioural analytics’ means different things to different people and is often banded about loosely. In many cases, it is used as a catch-all term for the analysis of ‘big data’ that is considered to represent ‘behaviour’ in some form or another, such as social media posts or website clicks, often with the purpose of developing a model that predicts, or seeks to influence, what someone does next.

Behavioural analytics emerged from business analytics and is often carried out by data or computer scientists without any training in behavioural science and with little to no knowledge of behavioural theory. For me the absence of behavioural science and theory in behavioural analytic models is problematic. Theory helps to guide the collection, processing and interpretation of the data, without which behavioural analytics is sometimes little more than a fishing exercise for discovering patterns in data. This is made doubly worse when the analytic methods used involve opaque machine learning algorithms that leave the analyst with no idea how or why a particular prediction was made.

To be fair, most computer and data-scientists are well aware of the risks of fishing for patterns and employ rigorous methods to avoid over-fitting models and reducing biases. But analytics in the absence of theory

still concerns me, particularly so when applied in the context of law enforcement and security. Why? There are several reasons; here are my top three:

DATA BLINKEREDNESS

Data is everywhere. Absolutely everywhere. With our smartphones and wearable technology, you can be firing signals about where you are (GPS), where you’re going and how you’re getting there (gmaps) and how much of a rush you’re in (heart-rate), all at the same time. With so much data trailing behind our every step, it is easy to forgive those who assume that everything we need to know must be lurking somewhere in the data, digitised into 0s and 1s, that just needs to be found. This can lead to too much focus on the data being used to detect and predict ‘behaviour’, with too little focus on the actual behaviour itself and how well the data represents this.

This is particularly the case when underestimating the importance of the context in which behaviour takes place. Real behaviour does not occur in a vacuum, and the fact you’ve just remembered you’ve left the oven on is not captured in the digital trace you leave behind as you rush home to turn it off.

SPURIOUS CORRELATIONS

With so much data, finding ‘significant’ patterns in the data is almost a statistical inevitability. Some of these patterns are likely to be meaningful, others less so. It can be all too easy to retrofit a convincing explanation onto a pattern that emerges to give it the air of legitimacy, but the reality is that often these rationales are guesswork. But if the model works, does this really matter? Yes,



it does. Retrofitted ‘rationales’ created in the vacuum of established theory may lead to an inaccurate understanding of reality. This may exacerbate prejudices and potentially even lead to ill-informed ‘interventions’. The age-old adage ‘correlation is not causation’ is never more important than when interpreting patterns identified in big data.

ADAPTING TO A CHANGING WORLD

Perhaps the biggest concern I have around atheoretical big data analytics is its reliance on modelling the world as it has been with an assumption that this is how it will be in the future. Of course, no one can build a model on data collected from the future, but a model that is based on a theoretical understanding of ‘why’ something occurs, rather than a collection of patterns, stands a better chance of adapting to unexpected shifts in the geo-socio-political context. It has been reported that the financial crisis of 2008 was partly caused by inadequate modelling of the factors that caused individuals to default on a loan. This led to a catastrophic under-estimation of the risk that something external to the model may cause a series of ‘unlikely’ events to happen at the same time. In the context of security, what terrorism in the UK has looked like over the past 20 years is very different to

“

The age-old adage ‘correlation is not causation’ is never more important than when interpreting patterns identified in big data.

”

what it looked like in the 20 years prior and might be very different to what it will look like in the next 20.

One could argue that a model built on a robust and established theory of what leads people to reject peace and accept violence as a legitimate tactic to achieve a political goal are perhaps more likely to detect the next wave of violent extremists than models built on patterns emerging from the swathes of data capturing who today’s terrorists are, where they live, or what beliefs they hold.