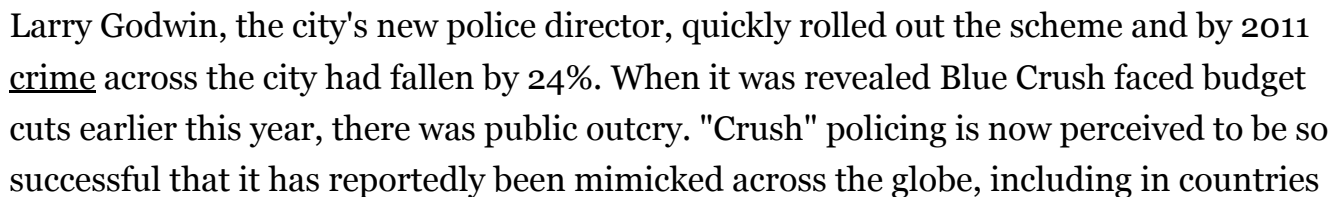
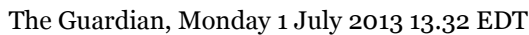


The NSA revelations highlight the role sophisticated algorithms play in sifting through masses of data. But more surprising is their widespread use in our everyday lives. So should we be more wary of their power?



such as Poland and Israel. In 2010, it was reported that two police forces in the UK were using it, but their identities were not revealed.

Crush stands for "Criminal Reduction Utilising Statistical History". Translated, it means predictive policing. Or, more accurately, police officers guided by algorithms. A team of criminologists and data scientists at the University of Memphis first developed the technique using IBM predictive analytics software. Put simply, they compiled crime statistics from across the city over time and overlaid it with other datasets – social housing maps, outside temperatures etc – then instructed algorithms to search for correlations in the data to identify crime "hot spots". The police then flooded those areas with highly targeted patrols.

"It's putting the right people in the right places on the right day at the right time," said Dr Richard Janikowski, an associate professor in the department of criminology and criminal justice at the University of Memphis, when the scheme launched. But not everyone is comfortable with the idea. Some critics have dubbed it "Minority Report" policing, in reference to the sci-fi film in which psychics are used to guide a "PreCrime" police unit.

The use of algorithms in policing is one example of their increasing influence on our lives. And, as their ubiquity spreads, so too does the debate around whether we should allow ourselves to become so reliant on them – and who, if anyone, is policing their use. Such concerns were sharpened further by the continuing revelations about how the US National Security Agency (NSA) has been using algorithms to help it interpret the colossal amounts of data it has collected from its covert dragnet of international telecommunications.

"For datasets the size of those the NSA collect, using algorithms is the only way to operate for certain tasks," says James Ball, the Guardian's data editor and part of the paper's NSA Files reporting team. "The problem is how the rules are set: it's impossible to do this perfectly. If you're, say, looking for terrorists, you're looking for something very rare. Set your rules too tight and you'll miss lots of, probably most, potential terror suspects. But set them more broadly and you'll drag lots of entirely blameless people into your dragnet, who will then face further intrusion or even formal investigation. We don't know exactly how the NSA or GCHQ use algorithms – or how extensively they're applied. But we do know they use them, including on the huge data trawls revealed in the Guardian."

From dating websites and City trading floors, through to online retailing and internet searches (Google's search algorithm is now a more closely guarded commercial secret

than the recipe for Coca-Cola), algorithms are increasingly determining our collective futures. "Bank approvals, store cards, job matches and more all run on similar principles," says Ball. "The algorithm is the god from the machine powering them all, for good or ill."



Most observers blame

the 'flash crash' of May 2010 on the use of algorithms to perform high-frequency trading. Photograph: Spencer Platt/Getty Images

But what is an algorithm? Dr Panos Parpas, a lecturer in the quantitative analysis and decision science ("quads") section of the department of computing at Imperial College London, says that wherever we use computers, we rely on algorithms: "There are lots of types, but algorithms, explained simply, follow a series of instructions to solve a problem. It's a bit like how a recipe helps you to bake a cake. Instead of having generic flour or a generic oven temperature, the algorithm will try a range of variations to produce the best cake possible from the options and permutations available."

Parpas stresses that algorithms are not a new phenomenon: "They've been used for decades – back to Alan Turing and the codebreakers, and beyond – but the current interest in them is due to the vast amounts of data now being generated and the need to process and understand it. They are now integrated into our lives. On the one hand, they are good because they free up our time and do mundane processes on our behalf. The questions being raised about algorithms at the moment are not about algorithms per se, but about the way society is structured with regard to data use and data privacy. It's also about how models are being used to predict the future. There is currently an awkward marriage between data and algorithms. As technology evolves, there will be mistakes, but it is important to remember they are just a tool. We shouldn't blame our tools."

The "mistakes" Parpas refers to are events such as the "flash crash" of 6 May 2010, when

the Dow Jones industrial average fell 1,000 points in just a few minutes, only to see the market regain itself 20 minutes later. The reasons for the sudden plummet has never been fully explained, but most financial observers blame a "race to the bottom" by the competing quantitative trading (quants) algorithms widely used to perform high-frequency trading. Scott Patterson, a Wall Street Journal reporter and author of The Quants, likens the use of algorithms on trading floors to flying a plane on autopilot. The vast majority of trades these days are performed by algorithms, but when things go wrong, as happened during the flash crash, humans can intervene.

"By far the most complicated algorithms are to be found in science, where they are used to design new drugs or model the climate," says Parpas. "But they are done within a controlled environment with clean data. It is easy to see if there is a bug in the algorithm. The difficulties come when they are used in the social sciences and financial trading, where there is less understanding of what the model and output should be, and where they are operating in a more dynamic environment. Scientists will take years to validate their algorithm, whereas a trader has just days to do so in a volatile environment."

Most investment banks now have a team of computer science PhDs coding algorithms, says Parpas, who used to work on such a team. "With City trading, everyone is running very similar algorithms," he says. "They all follow each other, meaning you get results such as the flash crash. They use them to speed up the process and to break up big trades to disguise them from competitors when a big investment is being made. It's an on-going, live process. They will run new algorithms for a few days to test them before letting them loose with real money. In currency trading, an algorithm lasts for about two weeks before it is stopped because it is surpassed by a new one. In equities, which is a less complicated market, they will run for a few months before a new one replaces them. It takes a day or two to write a currency algorithm. It's hard to find out information about them because, for understandable reasons, they don't like to advertise when they are successful. Goldman Sachs, though, has a strong reputation across the investment banks for having a brilliant team of algorithm scientists. PhDs students in this field will usually be employed within a few months by an investment bank."

The idea that the world's financial markets – and, hence, the wellbeing of our pensions, shareholdings, savings etc – are now largely determined by algorithmic vagaries is unsettling enough for some. But, as the NSA revelations exposed, the bigger questions surrounding algorithms centre on governance and privacy. How are they being used to access and interpret "our" data? And by whom?

Dr Ian Brown, the associate director of Oxford University's Cyber Security Centre, says

we all urgently need to consider the implications of allowing commercial interests and governments to use algorithms to analyse our habits: "Most of us assume that 'big data' is munificent. The laws in the US and UK say that much of this [the NSA revelations] is allowed, it's just that most people don't realise yet. But there is a big question about oversight. We now spend so much of our time online that we are creating huge data-mining opportunities."



Algorithms can run the risk of linking some racial groups to particular crimes. Photograph: Alamy Brown says that algorithms are now programmed to look for "indirect, non-obvious" correlations in data. "For example, in the US, healthcare companies can now make assessments about a good or bad insurance risk based, in part, on the distance you commute to work," he says. "They will identify the low-risk people and market their policies at them. Over time, this creates or exacerbates societal divides. Professor Oscar Gandy, at the University of Pennsylvania, has done research into 'secondary racial discrimination', whereby credit and health insurance, which relies greatly on postcodes, can discriminate against racial groups because they happen to live very close to other racial groups that score badly."

Brown harbours similar concerns over the use of algorithms to aid policing, as seen in Memphis where Crush's algorithms have reportedly linked some racial groups to particular crimes: "If you have a group that is disproportionately stopped by the police, such tactics could just magnify the perception they have of being targeted."

Viktor Mayer-Schönberger, professor of internet governance and regulation at the Oxford Internet Institute, also warns against humans seeing causation when an algorithm identifies a correlation in vast swaths of data. "This transformation presents an entirely new menace: penalties based on propensities," he writes in his new book, Big

Data: A Revolution That Will Transform How We Live, Work and Think, which is co-authored by Kenneth Cukier, the Economist's data editor. "That is the possibility of using big-data predictions about people to judge and punish them even before they've acted. Doing this negates ideas of fairness, justice and free will. In addition to privacy and propensity, there is a third danger. We risk falling victim to a dictatorship of data, whereby we fetishise the information, the output of our analyses, and end up misusing it. Handled responsibly, big data is a useful tool of rational decision-making. Wielded unwisely, it can become an instrument of the powerful, who may turn it into a source of repression, either by simply frustrating customers and employees or, worse, by harming citizens."

Mayer-Schönberger presents two very different real-life scenarios to illustrate how algorithms are being used. First, he explains how the analytics team working for US retailer Target can now calculate whether a woman is pregnant and, if so, when she is due to give birth: "They noticed that these women bought lots of unscented lotion at around the third month of pregnancy, and that a few weeks later they tended to purchase supplements such as magnesium, calcium and zinc. The team ultimately uncovered around two dozen products that, used as proxies, enabled the company to calculate a 'pregnancy prediction' score for every customer who paid with a credit card or used a loyalty card or mailed coupons. The correlations even let the retailer estimate the due date within a narrow range, so it could send relevant coupons for each stage of the pregnancy."

Harmless targeting, some might argue. But what happens, as has already reportedly occurred, when a father is mistakenly sent nappy discount vouchers instead of his teenage daughter whom a retailer has identified is pregnant before her own father knows?

Mayer-Schönberger's second example on the reliance upon algorithms throws up even more potential dilemmas and pitfalls: "Parole boards in more than half of all US states use predictions founded on data analysis as a factor in deciding whether to release somebody from prison or to keep him incarcerated."



Norah Jones: a

specially developed algorithm predicted that her debut album contained a disproportionately high number of hit records. Photograph: Olycom SPA/Rex Features Christopher Steiner, author of Automate This: How Algorithms Came to Rule Our World, has identified a wide range of instances where algorithms are being used to provide predictive insights – often within the creative industries. In his book, he tells the story of a website developer called Mike McCready, who has developed an algorithm to analyse and rate hit records. Using a technique called advanced spectral deconvolution, the algorithm breaks up each hit song into its component parts – melody, tempo, chord progression and so on – and then uses that to determine common characteristics across a range of No 1 records. McCready's algorithm correctly predicted – before they were even released – that the debut albums by both Norah Jones and Maroon 5 contained a disproportionately high number of hit records.

The next logical step – for profit-seeking record companies, perhaps – is to use algorithms to replace the human songwriter. But is that really an attractive proposition? "Algorithms are not yet writing pop music," says Steiner. He pauses, then laughs. "Not that we know of, anyway. If I were a record company executive or pop artist, I wouldn't tell anyone if I'd had a number one written by an algorithm."

Steiner argues that we should not automatically see algorithms as a malign influence on our lives, but we should debate their ubiquity and their wide range of uses. "We're already halfway towards a world where algorithms run nearly everything. As their power intensifies, wealth will concentrate towards them. They will ensure the 1%-99% divide gets larger. If you're not part of the class attached to algorithms, then you will struggle. The reason why there is no popular outrage about Wall Street being run by algorithms is because most people don't yet know or understand it."

But Steiner says we should welcome their use when they are used appropriately to aid and speed our lives. "Retail algorithms don't scare me," he says. "I find it useful when Amazon tells me what I might like. In the US, we know we will not have enough GP doctors in 15 years, as not enough are being trained. But algorithms can replace many of their tasks. Pharmacists are already seeing some of their prescribing tasks replaced by algorithms. Algorithms might actually start to create new, mundane jobs for humans. For example, algorithms will still need a human to collect blood and urine samples for them to analyse."

There can be a fine line, though, between "good" and "bad" algorithms, he adds: "I don't find the NSA revelations particularly scary. At the moment, they just hold the data. Even the best data scientists would struggle to know what to do with all that data. But it's the next step that we need to keep an eye on. They could really screw up someone's life with a false prediction about what they might be up to."

More from the Guardian

Working-class talent being priced out of acting, says David Morrissey

16 Sep 2014

Independent Scotland will face NHS budget cuts, says Alistair Darling

16 Sep 2014

French prime minister to seek vote of confidence from parliament

16 Sep 2014

Stonehenge researchers discover site is much larger than previously thought

10 Sep 2014

Spacewatch: Doubts about the Managua meteorite

11 Sep 2014

What's this?

More from around the web

Promoted content by Outbrain

How Much Will You Need to Save for the Retirement You Always Wanted

(Bring Your Challenges)

Your 401(k) Isn't Growing As Fast As it Should - Here's Why

(Mint)

Is This Tiny Stock the Next Big Thing?

(Venture Capital News)

How To Invest \$100, \$1,000, Or \$10,000

(Business Insider)

Warren Buffett's 10 Favorite Stocks

(Business Cheat Sheet)

Recommended by

© 2014 Guardian News and Media Limited or its affiliated companies. All rights reserved.

;