

# COSMOS

THE SCIENCE OF EVERYTHING

## THE END OF VIOLENCE

Steven Pinker on the new peace  
p36

## DEFEATING POLIO

Will politics jeopardise a cure? p68

## FRAUDS AND FAKES

Science's biggest scams p24

## GENIUS OF DOGS

Inside the canine brain p41

ISSUE 51 JUN/JUL 2013  
A\$10.95 NZ\$12.90



9 771832 522008  
COSMOSmagazine.com

# is data THE NEW GOD?

How tracking your digital trail could predetermine your future – and why you'll benefit from today's data deluge. p46

Preventing cybercrime



# The decision makers

Powerful computers are tracking our digital footprints, down to the most mundane activities in our day-to-day lives. **Gemma Black** looks at how personal information is shared – and whether you'll benefit from the new data deluge.



**I**N STEVEN SPIELBERG'S 2002 big-budget sci-fi film, *Minority Report*, a special police force exploits the skills of three genetically mutated humans with precognition powers, called 'precogs', to fight crime before it happens. At the time of the film's release, the notion of Tom Cruise barreling down on an unsuspecting future criminal who hadn't yet experienced the series of events that would lead to his or her crime – let alone the crime itself – seemed far-fetched at best, an affront to the presumption of innocence at worst, and challenged notions of free will and justice.

But just 10 years on, precogs exist. Not as mutated humans bathed in water and blue light, as imagined by Spielberg in his adaptation of Philip K. Dick's short story, but as powerful computers trawling

and an increased capacity to interpret and analyse that data, tied to an emerging awareness of its value and usefulness.

**WE'VE REACHED A ZENITH** of the digital age, in which the bulk of all information now available is digitised. That means it can be translated into the 0s and 1s needed for computer algorithms to test hypotheses against data correlations – and come up with answers to almost all our questions quickly and efficiently.

"All of the decisions people make in their daily lives are based on a very crude understanding of the world," says Tiberio Caetano, a computer scientist and machine learning expert at NICTA, Australia's information and communications technology research organisation.

"In the future, these decisions will be driven by data," Caetano continues. "There will be technology to help you be a happier person or make better decisions. It will not be a person, it will not be God, it will not be a friend – it will be computers." And this "paternalistic entity", as Caetano calls it, has the power to

**"There will be technology to help you be a happier person or make better decisions. It will not be a person, it will not be God, it will not be a friend – it will be computers."**

change the way we organise our lives. While NICTA is largely funded by the Australian Government, most of Caetano's current Big Data projects are with businesses trying to better understand their consumers. This can mean anything from the convenience of relevant reviews on retail websites, to the creepiness of department stores congratulating you on your pregnancy – without you telling them. More and more, retailers are looking for ways to mine this increasingly valuable resource.

A society, they claim, that is built around scientific principles of hypotheses and hard evidence. A society informed by what one computer scientist refers to as a "paternalistic entity" instead of the old-fashioned concept of opinion, and a society that will inevitably force a re-evaluation of our concept of privacy.

The catchphrase behind the concept is 'Big Data', which, to people in the field, means a lot more than simply 'lots of data'. Rather, Big Data – a sadly unimaginative marketing term that would probably make Spielberg cringe – refers to the relationship between the 21st-century data deluge

through oceans of raw data with the ability to learn from their findings as they go. Today's precogs face similar ethical hurdles to those in *Minority Report* – but also, enthusiasts insist, the potential to build a more efficient society.

According to computer scientist Rami Mukhtar, also at NICTA, being able to predict movements, buying habits and even fertility, isn't as difficult as one might think. "The simple hypothesis is that human beings are creatures of habit," he says. "The reality is, people don't do erratic things most of the time."



So, when a 15-year-old from Minneapolis, U.S., fit a 'pregnancy algorithm' at department store Target – by buying some of the 25 products that pregnant women often buy, including stretch-mark cream and vitamin supplements – the store sent her coupons for baby products. Her father went to his local store and demanded to see the manager, asking whether Target was encouraging pregnancy in high-school students. *The New York Times* reported in February 2012. But the story took a twist when he later found out that his daughter was indeed pregnant; the coupons were relevantly targeted, if tactless. The algorithms were accurate, and disconcertingly so (see 'Big Brother', below).

**THE METHODS** employed might sound simple enough. Eye anyone's trolley at the supermarket and you can probably get a pretty good idea of his or her lifestyle.

Information shared on social media is now a valuable commodity.

But multiply that by millions – and not just their shopping lists but their daily Tweets, Facebook status updates, phone calls, emails and Google searches – and clearly it's a task beyond any number of individuals to sift through and interpret that kind of data.

Instead, computers are taught what to look for and given the foundations from

**"These organisations have the potential to understand you much better than you do."**

which to work, then left to their own devices. It's called 'supervised machine learning' – present a computer with 100 images of a man and tell it they're men, and it will be able to tell you whether the never-before-seen 101st picture is a man, based on what it 'learned' from the previous lot.

"What we're doing is training a model, and what that model is doing is trying to recognise patterns

in all of those features which are highly correlated with a particular outcome, looking at all the historical samples with all of that information," explains Mukhtar.

"These organisations have the potential to understand you much better than you do," says Caetano.

In the case of the 15-year-old girl, while the department store may not have known more about her than she did, they certainly knew more than almost anyone else, including her family. What we're looking at, Caetano predicts, is a brave new world in which scientific principles are almost ubiquitous – for better or worse.

"Big Data, and the use of machine learning and all these algorithms and mathematics, is nothing other than exercising the philosophy of science in computers," he says.

"If you think of the history of science, it's all about posing hypotheses and guessing hypotheses. If you pose a hypothesis and

it's still surviving, you keep it. If not, you change the hypothesis," Caetano says. "We are doing that as we speak with computer algorithms – we are making them look at these oceans of data, then generate hypotheses and if those hypotheses are consistent with the data observed the next day, you keep the hypotheses. If not, you don't. Everything is becoming science. Every decision will be evidence-based."

**CONSIDER A** telecommunications company, previously at the mercy of both a fickle marketplace and forecasts by educated but limited marketing managers and chief financial officers. The company could supply an algorithm with all of the data it has about customers who cancelled their contracts in the past year, then use the same algorithm to analyse the data from the entire existing customer base. Uncovering correlations, the computer could predict which of those customers – to an astonishing degree of accuracy – is at risk of cancelling their contracts. That's when you might receive an enticing deal designed to persuade you to sign on for another year – conceivably, this could happen before you even realised you

were a prime candidate for becoming lost business.

Already, Big Data and machine learning permeates almost every aspect of our digital lives. Caetano says he recently translated some English text into his native language, Portuguese, using Google Translate to see how quickly the online tool was 'learning'. "It was doing so much better than three months ago," he says.

But for all the rhetoric of revolutions, there are still hurdles to overcome. Alan Dormer, a research leader in services science at Australia's national science agency, CSIRO, says sheer energy requirements alone will be a challenge. Data storage costs >>



Use of Big Data prompted a department store to market pregnancy products to a teenager, prompting an angry response from her family – despite the data's accuracy.

# Big Brother

ONE OF THE KEY, and most troubling, aspects of ambient intelligence is the constant connectivity brought about by our smartphones. There are more than a billion in use around the world, according to late-2012 stats from global analytics firm Strategy Analytics. For many, smartphones are more important than wallets and keys – they remind us of our itinerary, organise our events, and are increasingly being used to conduct financial transactions on the go.

As our phones leave their trail of digital footprints, the question becomes, how much information given out is too much? "The big challenge designers and implementers are facing is how to balance privacy and usability," says Daniel Cuthbert, Chief Operating Officer for information security company Sensepost. "How will these technologies be designed and secured so as not to put people at

risk but, at the same time, serve a purpose?"

Cuthbert's group recently discovered flaws in the security of smartphones as they try to connect to Wi-Fi providers. All smartphones can connect to Wi-Fi services, and most users leave this function switched on. "One of the flaws we found back in 2004 was that a mobile device will probe for the last x number of connected access points," Cuthbert says. These will include homes, cafes, shopping centres and other public hotspots we've logged onto through our day.

The team at Sensepost developed hacking software to identify what hotspots our phones are calling out for as they pass by. They were able to identify individual phones by their Media Access Control address, a unique identification code embedded in each

phone's hardware. By tracking a phone's Wi-Fi activity, hackers can track someone's movements, the researchers found. "We have built up a framework to figure out where people are, where they travel to, and where they've been," explains Cuthbert. "And in busy areas or busy times you can gather a lot of information in a short period."

Positioned at London Liverpool Street, one of London's busiest train stations, during rush hour in 2012, Cuthbert's team identified more than 11,000 phones. With his team then positioned at different stations around the city, they could have identified the movements of thousands over the course of a day.



The solution is to switch off your Wi-Fi when you're not using it. But most of us are now so used to being constantly connected that we may not want to – in effect, giving up privacy for convenience. Previous work by wireless security expert Don Bailey, CEO of San Francisco-based Capitol Hill Consultants revealed the ability of criminals to purchase mobile phone towers at cheap rates and tap into our phone calls. We are increasingly likely to be overheard and intercepted. Our reliance on technology, particularly smartphones, is also snowballing the trail of data we leave behind us. "If this single device has such access to your daily life, your movements and online behaviour, the security of the device has to be paramount," Cuthbert warns.

– Meera Senthilingam

# A happy home

AS YOU WAKE UP on a Monday morning, the lights in your pre-warmed house come on slowly, mimicking a natural sunrise. As you walk downstairs, the radio kicks in, and the coffee is already brewed. Later, the washing machine reminds you to add your laundry for it to wash and dry. You then leave for work, knowing that your lights and gadgets will switch off or on once your home knows you've gone.

Whoever said Monday morning was painful needed a world like this. It's a concept based on 'ambient intelligence', the idea that your smart surroundings can respond in ways it predicts you will find useful.

Michael O'Grady, an expert in mobile computing and intelligent information systems at University College Dublin, is one of many people working to bring this concept to reality. "The idea behind ambient intelligence is to provide seamless and intuitive interaction between you and your environment or technology," he says.

Diane Cook is another. A computer scientist at Washington State University in the U.S., she leads a team that has developed 30 smart home test beds – essentially

replicas of homes, apartments, offices and assisted living spaces equipped with sensors in the walls, and doorways and infrastructure that can monitor movement, temperature, light levels and water use. An overarching intelligent 'agent' collects information from the building's sensors and uses artificial intelligence software to process the information and pinpoint trends in the inhabitant's behaviour. The challenge is identifying the right mix of sensors and intelligence software to maximise the building's efficiency.

There's a variety of applications for the technology, says Cook. "For an energy-efficient goal, a home would be aware of a resident's activities and turn off devices that are not in use or not critical for an activity." A home that could monitor an individual's movements and behaviour would be ideal for sick or elderly people, with caregivers alerted if there is a deviation from someone's normal behaviour. For example, if a person falls over in the kitchen, sensors would register a lack of movement, setting off an alarm at a caregiver's home, or at a hospital. – Meera Senthilingam



Apart from the amount of data involved, the energy consumption of Big Data projects is a major logistical challenge.



may be plummeting, but in September 2011, Google revealed that its data centres consume 260 million watts of energy – enough to power Salt Lake City, Utah, with its 187,000-strong population.

"It may not be the case that you can't get enough disk drives, it may be that you

**What we're looking at, Caetano predicts, is a brave new world in which scientific principles are almost ubiquitous – for better or worse.**

can't afford to store it because of the cost of the electricity," Dormer says. "The data storage world is concerned about that."

Even without energy restraints, there's already too much data to store. The answer, Dormer predicts, will be "real-time situational awareness" – computers sorting

and interpreting data in real time, to be acted on as events occur. Data will simply be tossed out after it has been used.

Already, Dormer says, several Australian government departments monitor social media such as Facebook and Twitter in real time, with computer algorithms

analysing data to reveal spikes in certain terms, phrases and even sentiments, presented in simple visual format for continuous scrutinising.

"I think [Big Data] will enable a vast increase in efficiency," Dormer says. "People will stop doing things that aren't

worthwhile, because they will be able to see quite quickly if what they're doing will make a difference. Once we get this real-time situational awareness off the ground, it will enable super-rapid innovation."

Dormer says he is currently working on a project with the Indonesian government, in which regions vulnerable to poverty, unemployment or natural disasters are visualised on a sort of 'heat map' using available data to enable ongoing visualisation of the impacts of policies on those areas.

"The potential is that you [implement] a policy initiative and then you show a government official 'that was last week, now look at it'," explains Dormer. "The official can interpret that. He doesn't need an army of statisticians to tell him more green or less red is good."

Real-time analyses won't just look at the big-picture questions though. The key

could be in focussing on what most of us consider to be the mundane, such as the timing of your morning coffee.

"If I've stopped buying coffee, or if [I spend] more time in a café tomorrow, what does that say about my feelings of job security?" Mukhtar asks. And what does it mean for "the unemployment rate in six months time"? These types of questions, when paired with real-time analysis, could become more significant than your answers to the census.

The 2012 U.S. election was tied to Big Data, with reports that machine-learning algorithms were employed to target campaign messages to individual voters based on the available information about them. This included filtering voters by their 'persuadability', and then targeting those voters with campaign messages and policies most likely to sway their decision, as reported in *The Wall Street Journal* in November 2012.

"If you think about it, it has always been like this," says Caetano. That is, we've always collected data – in filing cabinets, with pen and paper, in our memories – and we've always used that data, or information, to inform our decisions. "It's just that it's going to happen at a much faster pace, and it's going to happen in a much more interesting way," he says.

**BUT ARE TODAY'S PRECOGS** really a social and marketing revolution? Or are they an invasion of privacy, placing too much power and knowledge in the hands of government and big corporations?

Target learned the hard way that if coupons and ads are too targeted, the company appears creepy and customers are turned off. Retailers these days have to appear as though they know less about their customers than they really do.

And each time Facebook updates its privacy or data-use policy, public trepidation ensues – not just from personal users, but also from governments. Most people are still wary of sacrificing privacy in the name of efficient marketing.

However, Caetano predicts this will change. Indeed, our relationship to privacy has already evolved and continues to do so – many of us upload information and photos about our lives, careers, relationships and weekend pursuits without hesitation. And after less than a decade, we already chuckle about the early days when it was common

to avoid using full names in email addresses to protect our identity.

"My hunch is that the understanding of the trade-off between the value of your data and the privacy of your data is going to evolve dramatically in the next 10 years," Caetano says.

"People will understand that to have a better service from this paternalistic entity, you need to disclose your data. There's no other way for this to happen. The price you pay is part, but not all, of what you've considered private so far," he says. "Obviously, each person should have the freedom to set the threshold of what they deem most appropriate," he adds.

"Like other socio-technical phenomena, Big Data triggers both utopian and dystopian rhetoric," wrote social media commentators Danah Boyd and Kate Crawford, from Microsoft Research in

Cambridge in the U.S., in a May 2012 paper, 'Critical Questions for Big Data', published in the journal *Information, Communication & Society*.

"There are some significant and insightful studies currently being done that involve Big Data, but it is still necessary to ask critical questions about what all this data means, who gets access to what data, how data analysis is deployed, and to what ends," they said.

Whether you call it Big Data, a paternalistic entity or Big Brother, the upshot seems to be that the comings and goings of our mundane lives have become more valuable than ever before. The consumer has, to a large extent, become the commodity – but it has to be up to us to decide whether it's worth the price. ■

Gemma Black is COSMOS's staff writer.

