About the Book

Would you let your family's full medical history be made public if it would help find a cure for cancer?

You are accused of a crime. Who do you want to decide your future – a human or a machine?

Which driverless car would you buy – one that prioritizes your life in the event of a crash, or one that reacts to save as many lives as possible?

Welcome to the age of the machine, the story of a not-too-distant future where algorithms rule supreme, making decisions about our healthcare, security, what we watch, where we go, even who goes to prison. So how much should we rely on them? What kind of future do we want?

Hannah Fry takes us on a tour of the good, the bad and the downright ugly of the algorithms that surround us. In *Hello World* she lifts the lid on their inner workings, demonstrates their power, exposes their limitations, and examines whether they really are an improvement on the humans they are replacing.

Contents

Cover
About the Book
Title Page
Dedication
A note on the title
Introduction

Power
Data
Justice
Medicine
Cars
Crime
Art

Conclusion
Photograph Credits
Notes
Acknowledgements
Index
About the Author
Also by Hannah Fry
Copyright

Hello world

How to Be Human in the Age of the Machine

HANNAH FRY

For Marie Fry.
Thank you for never taking no for an answer.

A note on the title

WHEN I WAS 7 years old, my dad brought a gift home for me and my sisters. It was a ZX Spectrum, a little 8-bit computer – the first time we'd ever had one of our own. It was probably already five years out of date by the time it arrived in our house, but even though it was second-hand, I instantly thought there was something marvellous about that dinky machine. The Spectrum was roughly equivalent to a Commodore 64 (although only the really posh kids in the neighbourhood had one of those) but I always thought it was a far more beautiful beast. The sleek black plastic casing could fit in your hands, and there was something rather friendly about the grey rubber keys and rainbow stripe running diagonally across one corner.

For me, the arrival of that ZX Spectrum marked the beginning of a memorable summer spent up in the loft with my elder sister, programming hangman puzzles for each other, or drawing simple shapes through code. All that 'advanced' stuff came later, though. First we had to master the basics.

Looking back, I don't exactly remember the moment I wrote my first ever computer program, but I'm pretty sure I know what it was. It would have been the same simple program that I've gone on to teach all of my students at University College London; the same as you'll find on the first page of practically any introductory computer science textbook. Because there is a tradition among all those who have ever learned to code – a rite of passage, almost. Your first task as a rookie is to program the computer to flash up a famous phrase on to the screen:

'HELLO WORLD'

It’s a tradition that dates back to the 1970s, when Brian Kernighan included it as a tutorial in his phenomenally popular programming textbook.[1] The book – and hence the phrase – marked an important point in the history of computers. The microprocessor had just arrived on the scene, heralding the transition of computers from what they had been in the past – enormous great specialist machines, fed on punch cards and ticker-tape – to something more like the personal computers we’re used to, with a screen, a keyboard and a blinking cursor. ‘Hello world’ came along at the first moment when chit-chat with your computer was a possibility.

Years later, Brian Kernighan told a *Forbes* interviewer about his inspiration for the phrase. He’d seen a cartoon showing an egg and a newly hatched chick chirping the words ‘Hello world!’ as it was born, and it had stuck in his mind.

It’s not entirely clear who the chick is supposed to be in that scenario: the fresh-faced human triumphantly announcing their brave arrival to the world of programming? Or the computer itself, awakening from the mundane slumber of spreadsheets and text documents, ready to connect its mind to the real world and do its new master’s bidding? Maybe both. But it’s certainly a phrase that unites all programmers, and connects them to every machine that’s ever been programmed.

There’s something else I like about the phrase – something that has never been more relevant or more important than it is now. As computer algorithms increasingly control and decide our future, ‘Hello world’ is a reminder of a moment of dialogue between human and machine. Of an instant where the boundary between controller and controlled is virtually imperceptible. It marks the start of a partnership – a shared journey of possibilities, where one cannot exist without the other.

In the age of the machine, that’s a sentiment worth bearing in mind.

Introduction

ANYONE WHO HAS ever visited Jones Beach on Long Island, New York, will have driven under a series of bridges on their way to the ocean. These bridges, primarily built to filter people on and off the highway, have an unusual feature. As they gently arc over the traffic, they hang extraordinarily low, sometimes leaving as little as 9 feet of clearance from the tarmac.

There's a reason for this strange design. In the 1920s, Robert Moses, a powerful New York urban planner, was keen to keep his newly finished, award-winning state park at Jones Beach the preserve of white and wealthy Americans. Knowing that his preferred clientele would travel to the beach in their private cars, while people from poor black neighbourhoods would get there by bus, he deliberately tried to limit access by building hundreds of low-lying bridges along the highway. Too low for the 12-foot buses to pass under.[1]

Racist bridges aren't the only inanimate objects that have had a quiet, clandestine control over people. History is littered with examples of objects and inventions with a power beyond their professed purpose.[2] Sometimes it's deliberately and maliciously factored into their design, but at other times it's a result of thoughtless omissions: just think of the lack of wheelchair access in some urban areas. Sometimes it's an unintended consequence, like the mechanized weaving machines of the nineteenth century. They were designed to make it easier to create complicated textiles, but in the end, the impact they had on wages, unemployment and working conditions made them arguably more tyrannical than any Victorian capitalist.

Modern inventions are no different. Just ask the residents of Scunthorpe in the north of England, who were blocked from opening AOL accounts after the internet giant created a new profanity filter that objected to the name of their town.[3] Or Chukwuemeka Afigbo, the Nigerian man who discovered an automatic hand-soap dispenser that released soap perfectly whenever his white friend placed their hand under the machine, but refused to acknowledge his darker skin.[4] Or Mark Zuckerberg, who, when writing the code for Facebook in his dorm room in Harvard in 2004, would never have imagined his creation would go on to be accused of helping manipulate votes in elections around the globe.[5]

Behind each of these inventions is an algorithm. The invisible pieces of code that form the gears and cogs of the modern machine age, algorithms have given the world everything from social media feeds to search engines, satellite navigation to music recommendation systems, and are as much a part of our modern infrastructure as bridges, buildings and factories ever were. They’re inside our hospitals, our courtrooms and our cars. They’re used by police forces, supermarkets and film studios. They have learned our likes and dislikes; they tell us what to watch, what to read and who to date. And all the while, they have the hidden power to slowly and subtly change the rules about what it means to be human.

In this book, we’ll discover the vast array of algorithms on which we increasingly, but perhaps unknowingly, rely. We’ll pay close attention to their claims, examine their undeclared power and confront the unanswered questions they raise. We’ll encounter algorithms used by police to decide who should be arrested, which make us choose between protecting the victims of crime and the innocence of the accused. We’ll meet algorithms used by judges to decide on the sentences of convicted criminals, which ask us to decide what our justice system should look like.

We'll find algorithms used by doctors to over-rule their own diagnoses; algorithms within driverless cars that insist we define our morality; algorithms that are weighing in on our expressions of emotion; and algorithms with the power to undermine our democracies.

I'm not arguing that algorithms are inherently bad. As you'll see in these pages, there are many reasons to be positive and optimistic about what lies ahead. No object or algorithm is ever either good or evil in itself. It's how they're used that matters. GPS was invented to launch nuclear missiles and now helps deliver pizzas. Pop music, played on repeat, has been deployed as a torture device. And however beautifully made a garland of flowers might be, if I really wanted to I could strangle you with it. Forming an opinion on an algorithm means understanding the relationship between human and machine. Each one is inextricably connected to the people who build and use it.

This means that, at its heart, this is a book about humans. It's about who we are, where we're going, what's important to us and how that is changing through technology. It's about our relationship with the algorithms that are already here, the ones working alongside us, amplifying our abilities, correcting our mistakes, solving our problems and creating new ones along the way.

It's about asking if an algorithm is having a net benefit on society. About when you should trust a machine over your own judgement, and when you should resist the temptation to leave machines in control. It's about breaking open the algorithms and finding their limits; and about looking hard at ourselves and finding our own. About separating the harm from the good and deciding what kind of world we want to live in.

Because the future doesn't just happen. We create it.

Power

GARRY KASPAROV KNEW exactly how to intimidate his rivals. At 34, he was the greatest chess player the world had ever seen, with a reputation fearsome enough to put any opponent on edge. Even so, there was one unnerving trick in particular that his competitors had come to dread. As they sat, sweating through what was probably the most difficult game of their life, the Russian would casually pick up his watch from where it had been lying beside the chessboard, and return it to his wrist. This was a signal that everybody recognized – it meant that Kasparov was bored with toying with his opponent. The watch was an instruction that it was time for his rival to resign the game. They could refuse, but either way, Kasparov's victory was soon inevitable.[1]

But when IBM's Deep Blue faced Kasparov in the famous match of May 1997, the machine was immune to such tactics. The outcome of the match is well known, but the story behind how Deep Blue secured its win is less widely appreciated. That symbolic victory, of machine over man, which in many ways marked the start of the algorithmic age, was down to far more than sheer raw computing power. In order to beat Kasparov, Deep Blue had to understand him not simply as a highly efficient processor of brilliant chess moves, but as a human being.

For a start, the IBM engineers made the brilliant decision to design Deep Blue to appear more uncertain than it was. During their infamous six-game match, the machine would occasionally hold off from declaring its move once a calculation had finished, sometimes for several minutes. From Kasparov's end of the table, the delays made it look as if the machine was struggling, churning through more and more calculations. It seemed to confirm what Kasparov thought he knew; that he'd successfully dragged the game into a position where the number of possibilities was so mind-bogglingly large that Deep Blue couldn't make a sensible decision.[2] In reality, however, it was sitting idly by, knowing exactly what to play, just letting the clock tick down. It was a mean trick, but it worked. Even in the first game of the match, Kasparov started to become distracted by second-guessing how capable the machine might be.[3]

Although Kasparov won the first game, it was in game two that Deep Blue really got into his head. Kasparov tried to lure the computer into a trap, tempting it to come in and capture some pieces, while at the same time setting himself up – several moves ahead – to release his queen and launch an attack.[4] Every watching chess expert expected the computer to take the bait, as did Kasparov himself. But somehow, Deep Blue smelt a rat. To Kasparov's amazement, the computer had realized what the grandmaster was planning and moved to block his queen, killing any chance of a human victory.[5]

Kasparov was visibly horrified. His misjudgement about what the computer could do had thrown him. In an interview a few days after the match he described Deep Blue as having 'suddenly played like a god for one moment'.[6] Many years later, reflecting on how he had felt at the time, he would write that he had 'made the mistake of assuming that moves that were surprising for a computer to make were also objectively strong moves'.[7] Either way, the genius of the algorithm had triumphed. Its understanding of the human mind, and human fallibility, was attacking and defeating the all-too-human genius.

Disheartened, Kasparov resigned the second game rather than fighting for the draw. From there his confidence began to unravel. Games three, four and five ended in draws. By game six, Kasparov was broken. The match ended Deep Blue 3½ to Kasparov's 2½.

It was a strange defeat. Kasparov was more than capable of working his way out of those positions on the board, but he had underestimated the ability of the algorithm and then allowed himself to be intimidated by it. 'I had been so impressed by Deep Blue's play,' he wrote in 2017, reflecting on the match. 'I became so concerned with what it might be capable of that I was oblivious to how my problems were more due to how badly I was playing than how well it was playing.'[8]

As we'll see time and time again in this book, expectations are important. The story of Deep Blue defeating the great grandmaster demonstrates that the power of an algorithm isn't limited to what is contained within its lines of code. Understanding our own flaws and weaknesses – as well as those of the machine – is the key to remaining in control.

But if someone like Kasparov failed to grasp this, what hope is there for the rest of us? Within these pages, we'll see how algorithms have crept into virtually every aspect of modern life – from health and crime to transport and politics. Along the way, we have somehow managed to be simultaneously dismissive of them, intimidated by them and in awe of their capabilities. The end result is that we have no idea quite how much power we're ceding, or if we've let things go too far.

Back to basics

Before we get to all that, perhaps it's worth pausing briefly to question what 'algorithm' actually means. It's a term that, although used frequently, routinely fails to convey much actual information. This is partly because the word itself is quite vague. Officially, it is defined as follows:[9]

> *algorithm* (noun): A step-by-step procedure for solving a problem or accomplishing some end especially by a computer.

That's it. An algorithm is simply a series of logical instructions that show, from start to finish, how to accomplish a task. By this broad definition, a cake recipe counts as an algorithm. So does a list of directions you might give to a lost stranger. IKEA manuals, YouTube troubleshooting videos, even self-help books – in theory, any self-contained list of instructions for achieving a specific, defined objective could be described as an algorithm.

But that's not quite how the term is used. Usually, algorithms refer to something a little more specific. They still boil down to a list of step-by-step instructions, but these algorithms are almost always mathematical objects. They take a sequence of mathematical operations – using equations, arithmetic, algebra, calculus, logic and probability – and translate them into computer code. They are fed with data from the real world, given an objective and set to work crunching through the calculations to achieve their aim. They are what makes computer science an actual science, and in the process have fuelled many of the most miraculous modern achievements made by machines.

There's an almost uncountable number of different algorithms. Each has its own goals, its own idiosyncrasies, its clever quirks and drawbacks, and there is no consensus on how best to group them. But broadly speaking, it can be useful to think of the real-world tasks they perform in four main categories:[10]

Prioritization: making an ordered list

Google Search predicts the page you're looking for by ranking one result over another. Netflix suggests which films you might like to watch next. Your TomTom selects your fastest route. All use a mathematical process to order the vast array of possible choices. Deep Blue was also essentially a prioritization algorithm, reviewing all the possible moves on the chessboard and calculating which would give the best chance of victory.

Classification: picking a category

As soon as I hit my late twenties, I was bombarded by adverts for diamond rings on Facebook. And once I eventually got married, adverts for pregnancy tests followed me around the internet. For these mild irritations, I had classification algorithms to thank. These algorithms, loved by advertisers, run behind the scenes and classify you as someone interested in those things on the basis of your characteristics. (They might be right, too, but it's still annoying when adverts for fertility kits pop up on your laptop in the middle of a meeting.)

There are algorithms that can automatically classify and remove inappropriate content on YouTube, algorithms that will label your holiday photos for you, and algorithms that can scan your handwriting and classify each mark on the page as a letter of the alphabet.

Association: finding links

Association is all about finding and marking relationships between things. Dating algorithms such as OKCupid have association at their core, looking for connections between members and suggesting matches based on the findings. Amazon's recommendation engine uses a similar idea, connecting your interests to those of past customers. It's what led to the intriguing shopping suggestion that confronted Reddit user Kerbobotat after buying a baseball bat on Amazon: 'Perhaps you'll be interested in this balaclava?'[11]

Filtering: isolating what's important

Algorithms often need to remove some information to focus on what's important, to separate the signal from the noise. Sometimes they do this literally: speech recognition algorithms, like those running inside Siri, Alexa and Cortana, first need to filter out your voice from the background noise before they can get to work on deciphering what you're saying. Sometimes they do it figuratively: Facebook and Twitter filter stories that relate to your known interests to design your own personalized feed.

The vast majority of algorithms will be built to perform a combination of the above. Take UberPool, for instance, which matches prospective passengers with others heading in the same direction. Given your start point and end point, it has to filter through the possible routes that could get you home, look for connections with other users headed in the same direction, and pick one group to assign you to – all while prioritizing routes with the fewest turns for the driver, to make the ride as efficient as possible.[12]

So, that's what algorithms can do. Now, how do they manage to do it? Well, again, while the possibilities are practically endless, there is a way to distil things. You can think of the approaches taken by algorithms as broadly fitting into two key paradigms, both of which we'll meet in this book.

Rule-based algorithms

The first type are rule-based. Their instructions are constructed by a human and are direct and unambiguous. You can imagine these algorithms as following the logic of a cake recipe. Step one: do this. Step two: if this, then that. That's not to imply that these algorithms are simple – there's plenty of room to build powerful programs within this paradigm.

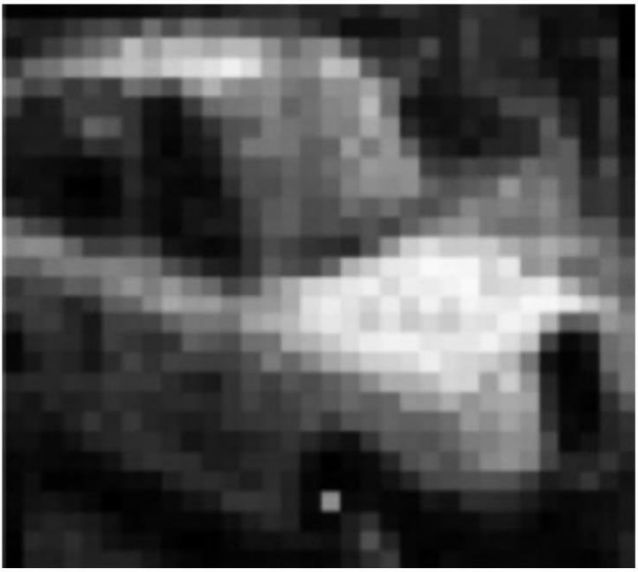
Machine-learning algorithms

The second type are inspired by how living creatures learn. To give you an analogy, think about how you might teach a dog to give you a high five. You don't need to produce a precise list of instructions and communicate them to the dog. As a trainer, all you need is a clear objective in your mind of what you want the dog to do and some way of rewarding her when she does the right thing. It's simply about reinforcing good behaviour, ignoring bad, and giving her enough practice to work out what to do for herself. The algorithmic equivalent is known as a *machine-learning algorithm*, which comes under the broader umbrella of *artificial intelligence* or *AI*. You give the machine data, a goal and feedback when it's on the right track – and leave it to work out the best way of achieving the end.

Both types have their pros and cons. Because rule-based algorithms have instructions written by humans, they're easy to comprehend. In theory, anyone can open them up and follow the logic of what's happening inside.[13] But their blessing is also their curse. Rule-based algorithms will only work for the problems for which humans know how to write instructions.

Machine-learning algorithms, by contrast, have recently proved to be remarkably good at tackling problems where writing a list of instructions won't work. They can recognize objects in pictures, understand words as we speak them and translate from one language to another – something rule-based algorithms have always struggled with. The downside is that if you let a machine figure out the solution for itself, the route it takes to get there often won't make a lot of sense to a human observer. The insides can be a mystery, even to the smartest of living programmers.

Take, for instance, the job of image recognition. A group of Japanese researchers recently demonstrated how strange an algorithm's way of looking at the world can seem to a human. You might have come across the optical illusion where you can't quite tell if you're looking at a picture of a vase or of two faces (if not, there's an example in the notes at the back of the book).[14] Here's the computer equivalent. The team showed that changing a single pixel on the front wheel of the image overleaf was enough to cause a machine-learning algorithm to change its mind from thinking this is a photo of a car to thinking it is a photo of a dog.[15]



For some, the idea of an algorithm working without explicit instructions is a recipe for disaster. How can we control something we don't understand? What if the capabilities of sentient, super-intelligent machines transcend those of their makers? How will we ensure that an AI we don't understand and can't control isn't working against us?

These are all interesting hypothetical questions, and there is no shortage of books dedicated to the impending threat of an AI apocalypse. Apologies if that was what you were hoping for, but this book isn't one of them. Although AI has come on in leaps and bounds of late, it is still only 'intelligent' in the narrowest sense of the word. It would probably be more useful to think of what we've been through as a revolution in computational statistics than a revolution in intelligence. I know that makes it sound a lot less sexy (unless you're *really* into statistics), but it's a far more accurate description of how things currently stand.

For the time being, worrying about evil AI is a bit like worrying about overcrowding on Mars.[fn1] Maybe one day we'll get to the point where computer intelligence surpasses human intelligence, but we're nowhere near it yet. Frankly, we're still quite a long way away from creating hedgehog-level intelligence. So far, no one's even managed to get past worm.[fn2]

Besides, all the hype over AI is a distraction from much more pressing concerns and – I think – much more interesting stories. Forget about omnipotent artificially intelligent machines for a moment and turn your thoughts from the far distant future to the here and now – because there are already algorithms with free rein to act as autonomous decision-makers. To decide prison terms, treatments for cancer patients and what to do in a car crash. They're already making life-changing choices on our behalf at every turn.

The question is, if we're handing over all that power – are they deserving of our trust?

Blind faith

Sunday, 22 March 2009 wasn't a good day for Robert Jones. He had just visited some friends and was driving back through the pretty town of Todmorden in West Yorkshire when he noticed the fuel light on his BMW. He had just 7 miles to find a petrol station before he ran out, which was cutting things rather fine. Thankfully his GPS seemed to have found him a short cut – sending him on a narrow winding path up the side of the valley.

Robert followed the machine's instructions, but as he drove, the road got steeper and narrower. After a couple of miles, it turned into a dirt track that barely seemed designed to accommodate horses, let alone cars. But Robert wasn't fazed. He drove five thousand miles a week for a living and knew how to handle himself behind the wheel. Plus, he thought, he had 'no reason not to trust the TomTom sat-nav'.[16]

Just a short while later, anyone who happened to be looking up from the valley below would have seen the nose of Robert's BMW appearing over the brink of the cliff above, saved from the hundred-foot drop only by the flimsy wooden fence at the edge he'd just crashed into.

It would eventually take a tractor and three quad bikes to recover Robert's car from where he abandoned it. Later that year, when he appeared in court on charges of reckless driving, he admitted that he didn't think to over-rule the machine's instructions. 'It kept insisting the path was a road,' he told a newspaper after the incident. 'So I just trusted it. You don't expect to be taken nearly over a cliff.'[17]

No, Robert. I guess you don't.

There's a moral somewhere in this story. Although he probably felt a little foolish at the time, in ignoring the information in front of his eyes (like seeing a sheer drop out of the car window) and attributing greater intelligence to an algorithm than it deserved, Jones was in good company. After all, Kasparov had fallen into the same trap some twelve years earlier. And, in much quieter but no less profound ways, it's a mistake almost all of us are guilty of making, perhaps without even realizing.

Back in 2015 scientists set out to examine how search engines like Google have the power to alter our view of the world.[18] They wanted to find out if we have healthy limits in the faith we place in their results, or if we would happily follow them over the edge of a metaphorical cliff.

The experiment focused around an upcoming election in India. The researchers, led by psychologist Robert Epstein, recruited 2,150 undecided voters from around the country and gave them access to a specially made search engine, called 'Kadoodle', to help them learn more about the candidates before deciding who they would vote for.

Kadoodle was rigged. Unbeknown to the participants, they had been split into groups, each of which was shown a slightly different version of the search engine results, biased towards one candidate or another. When members of one group visited the website, all the links at the top of the page would favour one candidate in particular, meaning they'd have to scroll right down through link after link before finally finding a single page that was favourable to anyone else. Different groups were nudged towards different candidates.

It will come as no surprise that the participants spent most of their time reading the websites flagged up at the top of the first page – as that old internet joke says, the best place to hide a dead body is on the second page of Google search results. Hardly anyone in the experiment paid much attention to the links that appeared well down the list. But still, the degree to which the ordering influenced the volunteers' opinions shocked even Epstein. After only a few minutes of looking at the search engine's biased results, when asked who they would vote for, participants were a staggering 12 per cent more likely to pick the candidate Kadoodle had favoured.

In an interview with *Science* in 2015,[19] Epstein explained what was going on: 'We expect the search engine to be making wise choices. What they're saying is, "Well yes, I see the bias and that's telling me ... the search engine is doing its job."' Perhaps more ominous, given how much of our information we now get from algorithms like search engines, is how much agency people believed they had in their own opinions: 'When people are unaware they are being manipulated, they tend to believe they have adopted their new thinking voluntarily,' Epstein wrote in the original paper.[20]

Kadoodle, of course, is not the only algorithm to have been accused of subtly manipulating people's political opinions. We'll come on to that more in the 'Data' chapter, but for now it's worth noting how the experiment suggests we feel about algorithms that are right most of the time. We end up believing that they always

> It's just this bias we all have for computerized results – we don't question them. When a computer generates something – when you have a statistician, who looks at some data, and comes up with a formula – we just trust that formula, without asking 'hey wait a second, how is this actually working?'[33]

Now, I know that picking mathematical formulae apart to see how they work isn't everyone's favourite pastime (even if it is mine). But Eppink none the less raises an incredibly important point about our human willingness to take algorithms at face value without wondering what's going on behind the scenes.

In my years working as a mathematician with data and algorithms, I've come to believe that the only way to objectively judge whether an algorithm is trustworthy is by getting to the bottom of how it works. In my experience, algorithms are a lot like magical illusions. At first they appear to be nothing short of actual wizardry, but as soon as you know how the trick is done, the mystery evaporates. Often there's something laughably simple (or worryingly reckless) hiding behind the façade. So, in the chapters that follow, and the algorithms we'll explore, I'll try to give you a flavour of what's going on behind the scenes where I can. Enough to see how the tricks are done – even if not quite enough to perform them yourself.

But even for the most diehard maths fans, there are still going to be occasions where algorithms demand you take a blind leap of faith. Perhaps because, as with Skyscanner or Google's search results, double-checking their working isn't feasible. Or maybe, like the Idaho budget tool and others we'll meet, the algorithm is considered a 'trade secret'. Or perhaps, as in some machine-learning techniques, following the logical process inside the algorithm just isn't possible.

There will be times when we have to hand over control to the unknown, even while knowing that the algorithm is capable of making mistakes. Times when we are forced to weigh up our own judgement against that of the machine. When, if we decide to trust our instincts instead of its calculations, we're going to need rather a lot of courage in our convictions.

When to over-rule

Stanislav Petrov was a Russian military officer in charge of monitoring the nuclear early warning system protecting Soviet airspace. His job was to alert his superiors immediately if the computer indicated any sign of an American attack.[34]

Petrov was on duty on 26 September 1983 when, shortly after midnight, the sirens began to howl. This was the alert that everyone dreaded. Soviet satellites had detected an enemy missile headed for Russian territory. This was the depths of the Cold War, so a strike was certainly plausible, but something gave Petrov pause. He wasn't sure he trusted the algorithm. It had only detected five missiles, which seemed like an illogically small opening salvo for an American attack.[35]

Petrov froze in his chair. It was down to him: report the alert, and send the world into almost certain nuclear war; or wait, ignoring protocol, knowing that with every second that passed his country's leaders had less time to launch a counter-strike.

Fortunately for all of us, Petrov chose the latter. He had no way of knowing for sure that the alarm had sounded in error, but after 23 minutes – which must have felt like an eternity at the time – when it was clear that no nuclear missiles had landed on Russian soil, he finally knew that he had been correct. The algorithm had made a mistake.

If the system had been acting entirely autonomously, without a human like Petrov to act as the final arbiter, history would undoubtedly have played out rather differently. Russia would almost certainly have launched what it believed to be retaliatory action and triggered a full-blown nuclear war in the process. If there's anything we can learn from this story, it's that the human element does seem to be a critical part of the process: that having a person with the power of veto in a position to review the suggestions of an algorithm before a decision is made is the only sensible way to avoid mistakes.

After all, only humans will feel the weight of responsibility for their decisions. An algorithm tasked with communicating up to the Kremlin wouldn't have thought for a second about the potential ramifications of such a decision. But Petrov, on the other hand? 'I knew perfectly well that nobody would be able to correct my mistake if I had made one.'[36]

The only problem with this conclusion is that humans aren't always that reliable either. Sometimes, like Petrov, they'll be right to over-rule an algorithm. But often our instincts are best ignored.

To give you another example from the world of safety, where stories of humans incorrectly over-ruling an algorithm are mercifully rare, that is none the less precisely what happened during an infamous crash on the Smiler rollercoaster at Alton Towers, the UK's biggest theme park.[37]

Back in June 2015, two engineers were called to attend a fault on a rollercoaster. After fixing the issue, they sent an empty carriage around to test everything was working – but failed to notice it never made it back. For whatever reason, the spare carriage rolled backwards down an incline and came to a halt in the middle of the track.

Meanwhile, unbeknown to the engineers, the ride staff added an extra carriage to deal with the lengthening queues. Once they got the all-clear from the control room, they started loading up the carriages with cheerful passengers, strapping them in and sending the first car off around the track, completely unaware of the empty, stranded carriage sent out by the engineers sitting directly in its path.

Luckily, the rollercoaster designers had planned for a situation like this, and their safety algorithms worked exactly as planned. To avoid a certain collision, the packed train was brought to a halt at the top of the first climb, setting off an alarm in the control room. But the engineers – confident that they'd just fixed the ride – concluded the automatic warning system was at fault.

Over-ruling the algorithm wasn't easy: they both had to agree and simultaneously press a button to restart the rollercoaster. Doing so sent the train full of people over the drop to crash straight into the stranded extra carriage. The result was horrendous. Several people suffered devastating injuries and two teenage girls lost their legs.

Both of these life-or-death scenarios, Alton Towers and Petrov's alarm, serve as dramatic illustrations of a much deeper dilemma. In the balance of power between human and algorithm, who – or what – should have the final say?

Power struggle

This is a debate with a long history. In 1954, Paul Meehl, a professor of clinical psychology at the University of Minnesota, annoyed an entire generation of humans when he published *Clinical versus Statistical Prediction*, coming down firmly on one side of the argument.[38]

In his book, Meehl systematically compared the performance of humans and algorithms on a whole variety of subjects – predicting everything from students' grades to patients' mental health outcomes – and concluded that mathematical algorithms, no matter how simple, will almost always make better predictions than people.

Countless other studies in the half-century since have confirmed Meehl's findings. If your task involves any kind of calculation, put your money on the algorithm every time: in making medical diagnoses or sales forecasts, predicting suicide attempts or career satisfaction, and assessing everything from fitness for military service to projected academic performance.[39] The machine won't be perfect, but giving a human a veto over the algorithm would just add more error.[fn3]

Perhaps this shouldn't come as a surprise. We're not built to compute. We don't go to the supermarket to find a row of cashiers eyeballing our shopping to gauge how much it should cost. We get an (incredibly simple) algorithm to calculate it for us instead. And most of the time, we'd be better off leaving the machine to it. It's like the saying among airline pilots that the best flying team has three components: a pilot, a computer and a dog. The computer is there to fly the plane, the pilot is there to feed the dog. And the dog is there to bite the human if it tries to touch the computer.

But there's a paradox in our relationship with machines. While we have a tendency to over-trust anything we don't understand, as soon as we *know* an algorithm can make mistakes, we also have a rather annoying habit of over-reacting and dismissing it completely, reverting instead to our own flawed judgement. It's known to researchers as *algorithm aversion*. People are less tolerant of an algorithm's mistakes than of their own – even if their own mistakes are bigger.

It's a phenomenon that has been demonstrated time and time again in experiments,[40] and to some extent, you might recognize it in yourself. Whenever Citymapper says my journey will take longer than I expect it to, I always think I know better (even if most of the time it means I end up arriving late). We've all called Siri an idiot at least once, somehow in the process forgetting the staggering technological accomplishment it has taken to build a talking assistant you can hold in your hand. And in the early days of using the mobile GPS app Waze I'd found myself sitting in a traffic jam, having been convinced that taking the back roads would be faster than the route shown. (It almost always wasn't.) Now I've come to trust it and – like Robert Jones and his BMW – I'll blindly follow it wherever it leads me (although I still think I'd draw the line at going over a cliff).

had the upper hand. The data revealed which customers came back day after day, and which saved their shopping for weekends. Armed with that knowledge, they could get to work nudging their customers' buying behaviour, by sending out a series of coupons to the Clubcard users in the post. High spenders were given vouchers ranging from £3 to £30. Low spenders were sent a smaller incentive of £1 to £10. And the results were staggering. Nearly 70 per cent of the coupons were redeemed, and while in the stores, customers filled up their baskets: people who had Clubcards spent 4 per cent more overall than those who didn't.

On 22 November 1994, Clive Humby presented the findings from the trial to the Tesco board. He showed them the data, the response rates, the evidence of customer satisfaction, the sales boosts. The board listened in silence. At the end of the presentation, the chair was the first person to speak. 'What scares me about this,' he said, 'is that you know more about my customers in three months than I know in 30 years.'[3]

Clubcard was rolled out to all customers of Tesco and is widely credited with putting the company ahead of its main rival Sainsbury's, to become the biggest supermarket in the UK. As time wore on, the data collected became more detailed, making customers' buying habits easier to target.

Early in the days of online shopping, the team introduced a feature known as 'My Favourites', in which any items that were bought while using the loyalty card would appear prominently when the customer logged on to the Tesco website. Like the Clubcard itself, the feature was a roaring success. People could quickly find the products they wanted without having to navigate through the various pages. Sales went up, customers were happy.

But not all of them. Shortly after the launch of the feature, one woman contacted Tesco to complain that her data was wrong. She'd been shopping online and seen condoms among her list of 'My Favourites'. They couldn't be her husband's, she explained, because he didn't use them. At her request, the Tesco analysts looked into the data and discovered that her list was accurate. However, rather than be the cause of a marital rift, they took the diplomatic decision to apologize for 'corrupted data' and remove the offending items from her favourites.

According to Clive Humby's book on Tesco, this has now become an informal policy within the company. Whenever something comes up that is just a bit too revealing, they apologize and delete the data. It's a stance that's echoed by Eric Schmidt, who, while serving as the executive chairman of Google, said he tries to think of things in terms of an imaginary creepy line. 'The Google policy is to get right up to the creepy line but not cross it.'[4]

But collect enough data and it's hard to know what you'll uncover. Groceries aren't just what you consume. They're personal. Look carefully enough at someone's shopping habits and they'll often reveal all kinds of detail about who they are as a person. Sometimes – as in the case of the condoms – it'll be things you'd rather not know. But more often than not, lurking deep within the data, those slivers of hidden insight can be used to a company's advantage.

Target market

Back in 2002, the American discount superstore Target started looking for unusual patterns in its data.[5] Target sells everything from milk and bananas to cuddly toys and garden furniture, and – like pretty much every other retailer since the turn of the millennium – has ways of using credit card numbers and survey responses to tie customers to everything they've ever bought in the store, enabling them to analyse what people are buying.

In a story that – as US readers won't need telling – became infamous across the country, Target realized that a spike in a female customer's purchases of unscented body lotion would often precede her signing up to the in-store baby-shower registry. It had found a signal in the data. As women entered their second trimester and started to worry about stretch marks, their buying of moisturizer to keep their skin supple left a hint of what was to come. Scroll backwards further in time, and these same women would be popping into Target to stock up on various vitamins and supplements, like calcium and zinc. Scroll forwards in time and the data would even suggest when the baby was due – marked by the woman buying extra-big bags of cotton wool from the store.[6]

Expectant mothers are a retailer's dream. Lock in her loyalty while she's pregnant and there's a good chance she'll continue to use your products long after the birth of her child. After all, shopping habits are quick to form when a hungry screaming baby is demanding your attention during your weekly shop. Insights like this could be hugely valuable in giving Target a head start over other brands in attracting her business.

From there it was simple. Target ran an algorithm that would score its female customers on the likelihood they were pregnant. If that probability tipped past a certain threshold, the retailer would automatically send out a series of coupons to the woman in question, full of things she might find useful: nappies, lotions, baby wipes and so on.

So far, so uncontroversial. But then, around a year after the tool was first introduced, a father of a teenage girl stormed into a Target store in Minneapolis demanding to see the manager. His daughter had been sent some pregnancy coupons in the post and he was outraged that the retailer seemed to be normalizing teenage pregnancy. The manager of the store apologized profusely and called the man's home a few days later to reiterate the company's regret about the whole affair. But by then, according to a story in the *New York Times*, the father had an apology of his own to make.

'I had a talk with my daughter,' he told the manager. 'It turns out there's been some activities in my house I haven't been completely aware of. She's due in August.'

I don't know about you, but for me, an algorithm that will inform a parent that their daughter is pregnant before they've had a chance to learn about it in person takes a big step across the creepy line. But this embarrassment wasn't enough to persuade Target to scrap the tool altogether.

A Target executive explained: 'We found out that as long as a pregnant woman thinks she hasn't been spied on, she'll use the coupons. She just assumes that everyone else on her block got the same mailer for diapers and cribs. As long as we don't spook her, it works.'

So, Target still has a pregnancy predictor running behind the scenes – as most retailers do now. The only difference is that it will mix in the pregnancy-related coupons with other more generic items so that the customers don't notice they've been targeted. An advertisement for a crib might appear opposite some wine glasses. Or a coupon for baby clothes will run alongside an ad for some cologne.

Target is not alone in using these methods. Stories of what can be inferred from your data rarely hit the press, but the algorithms are out there, quietly hiding behind the corporate front lines. About a year ago, I got chatting to a chief data officer of a company that sells insurance. They had access to the full detail of people's shopping habits via a supermarket loyalty scheme. In their analysis, they'd discovered that home cooks were less likely to claim on their home insurance, and were therefore more profitable. It's a finding that makes good intuitive sense. There probably isn't much crossover between the group of people who are willing to invest time, effort and money in creating an elaborate dish from scratch and the group who would let their children play football in the house. But how did they know which shoppers were home cooks? Well, there were a few items in someone's basket that were linked to low claim rates. The most significant, he told me, the one that gives you away as a responsible, houseproud person more than any other, was fresh fennel.

If that's what you can infer from people's shopping habits in the physical world, just imagine what you might be able to infer if you had access to more data. Imagine how much you could learn about someone if you had a record of everything they did online.

The Wild West

Palantir Technologies is one of the most successful Silicon Valley start-ups of all time. It was founded in 2003 by Peter Thiel (of PayPal fame), and at the last count was estimated to be worth a staggering

$20 billion.[7] That's about the same market value as Twitter, although chances are you've never heard of it. And yet – trust me when I tell you – Palantir has most certainly heard of you.

Palantir is just one example of a new breed of companies whose business is our data. And alongside the analysts, there are also the data brokers: companies who buy and collect people's personal information and then resell it or share it for profit. Acxiom, Corelogic, Datalogix, eBureau – a swathe of huge companies you've probably never directly interacted with, that are none the less continually monitoring and analysing your behaviour.[8]

Every time you shop online, every time you sign up for a newsletter, or register on a website, or enquire about a new car, or fill out a warranty card, or buy a new home, or register to vote – every time you hand over any data at all – your information is being collected and sold to a data broker. Remember when you told an estate agent what kind of property you were looking for? Sold to a data broker. Or those details you once typed into an insurance comparison website? Sold to a data broker. In some cases, even your entire browser history can be bundled up and sold on.[9]

It's the broker's job to combine all of that data, cross-referencing the different pieces of information they've bought and acquired, and then create a single detailed file on you: a data profile of your digital shadow. In the most literal sense, within some of these brokers' databases, you could open up a digital file with your ID number on it (an ID you'll never be told) that contains traces of everything you've ever done. Your name, your date of birth, your religious affiliation, your vacation habits, your credit-card usage, your net worth, your weight, your height, your political affiliation, your gambling habits, your disabilities, the medication you use, whether you've had an abortion, whether your parents are divorced, whether you're easily addictable, whether you are a rape victim, your opinions on gun control, your projected sexual orientation, your real sexual orientation, and your gullibility. There are thousands and thousands of details within thousands and thousands of categories and files stored on hidden servers somewhere, for virtually every single one of us.[10]

Like Target's pregnancy predictions, much of this data is inferred. A subscription to *Wired* magazine might imply that you're interested in technology; a firearms licence might imply that you're interested in hunting. All along the way, the brokers are using clever, but simple, algorithms to enrich their data. It's exactly what the supermarkets were doing, but on a massive scale.

And there are plenty of benefits to be had. Data brokers use their understanding of who we are to prevent fraudsters from impersonating unsuspecting consumers. Likewise, knowing our likes and dislikes means that the adverts we're served as we wander around the internet are as relevant to our interests and needs as possible. That almost certainly makes for a more pleasant experience than being hit with mass market adverts for injury lawyers or PPI claims day after day. Plus, because the messages can be directly targeted on the right consumers, it means advertising is cheaper overall, so small businesses with great products can reach new audiences, something that's good for everyone.

But, as I'm sure you're already thinking, there's also an array of problems that arise once you start distilling who we are as people down into a series of categories. I'll get on to that in a moment, but first I think it's worth briefly explaining the invisible process behind how an online advert reaches you when you're clicking around on the internet, and the role that a data broker plays in the process.

So, let's imagine I own a luxury travel company, imaginatively called Fry's. Over the years, I have been getting people to register their interest on my website and now have a list of their email addresses. If I wanted to find out more about my users – like what kind of holidays they were interested in – I could send off my list of users' emails to a data broker, who would look up the names in their system, and return my list with the relevant data attached. Sort of like adding an extra column on to a spreadsheet. Now when you visit my Fry's website, I can see that you have a particular penchant for tropical islands and so serve you up an advert for a Hawaii getaway.

That's option one. In option two, let's imagine that Fry's has a little extra space on its website that we're willing to sell to other advertisers. Again, I contact a data broker and give them the information I