EVIL COMPUTERS

THE DARK SIDE OF COMPUTERS

Technology may have enriched all our lives, but it has also been used for nefarious purposes by unscrupulous governments, criminals and terrorists. Simon Handby reveals how computers have helped to perpetrate evil around the world

During the six decades or so that they've been around, computers have brought dramatic changes to our understanding of the world and the ways in which we work and communicate. The technology has penetrated our homes, cars and pockets, driving changes in our culture, behaviour and governance.

Most of this change has been for the good: we now communicate more easily, can work more efficiently and are better entertained than ever before. However, the foundations of today's technology were built amid the darker motives and necessities of the Second World War and the Cold War that followed, and the history of computing contains troubling examples of the pursuit of power and profit at the expense of people's lives.

Computers have been used for evil deeds, and through bugs or negligence have accidentally committed dreadful acts, but the ongoing development of artificial intelligence and autonomous systems raises an even more frightening prospect. Could ever more intelligent computers be used for ever greater evil, or could they leap above the humanity that created them and, living up to the darkest imaginings of science fiction, themselves become evil? Could technological evolution reach a tipping point beyond which humans, when it comes to survival, are no longer 'the fittest'?

GOING BALLISTIC

Computer-based technologies such as GPS and digital X-rays help to protect and save lives everyday, but modern computers are built upon advances undertaken in darker times. The origins of computing were innocent enough, with some of the earliest programmable machines developed by Joseph Marie Jacquard to automate looms in the textile industry in the 1800s. The first theoretical computer, Charles Babbage's 'analytical engine', was originally devised simply to remove human errors from the mathematical tables available in the early 19th century.

The motivations for these inventions may have been innocent, but the computer as we understand it today wasn't fully imagined until the years leading up to World War II, and it was the war that provided the money, facilities and impetus for the theories of computer scientists such as Alan Turing to be made real. It was the need among the analysts at Bletchley Park for massive computing power that drove development, first of the electromechanical 'Bombe' and 'Heath Robinson' machines and subsequently of Colossus – the first programmable, digital, electronic computer.

Being the product of a war effort doesn't automatically render a computer 'evil', of course. The Colossus computers were famously

→

EVIL COMPUTERS

used to mount 'brute force' attacks on the Lorenz cipher used by the German High Command, work that helped to save Allied lives and almost certainly shortened the duration of the war in Europe, even if Axis soldiers, and inevitably civilians, were killed in actions taken on the basis of the intelligence. Things are muddier, though, for other war-era computers, such as ENIAC.

ENIAC, a huge electronic 'brain' weighing more than 27 tonnes and containing more than 17,000 thermionic valves, was commissioned and funded by the United States Army and developed in secret at the University of Pennsylvania from 1943. Operational from 1946 until 1955, ENIAC was a ballistics computer, designed specifically to calculate artillery firing tables.

The very reason for its existence was to improve the accuracy and deadliness of the army's firepower, but while still under development it came to the attention of the mathematician John von Neumann, then working in the Manhattan Project on the development of the hydrogen bomb. The computer's first test run was computations for the bomb. It's hard to say whether ENIAC was instrumental, but the lethality of mankind's arsenal has certainly been improved thanks to computers.

↑ The Third Reich used punchcards to classify and tabulate the religion and sexuality of those it persecuted

DARK HISTORY

While the building of weapons (nuclear or otherwise) is a divisive debate, often dependent on who and why they are deployed, other examples of computer use seem harder to defend. In the 2001 book *IBM and the Holocaust*, US journalist Edwin Black alleges that IBM and its German subsidiary Dehomag developed and continued business relationships with the Nazi regime from Hitler's 1933 rise to power until the 1945 downfall of the Third Reich.

In particular, Black looks at the role of Hollerith punchcard machines, supplied by Dehomag, in the identification and cataloguing of Jews in the 1930s, and of IBM technology in the organisation of railroads and registration at concentration camps. His book alleges that IBM's subsidiaries leased, rather than sold, equipment to the Third Reich, that they maintained and upgraded punchcard machines throughout the war, and that Dehomag trained Nazi officers including concentration camp administrators.

The full extent of IBM's involvement with the Third Reich is disputed, but Dehomag, which had come under the control of Nazi authorities, did provide Hollerith equipment that was used for census operations vital to the Third Reich as it pursued various actions against its own citizens and those of annexed and invaded countries. Although the systems, designed to log data read from punchcards, weren't strictly computers, they were close cousins, and there's no doubt they helped to make terrible acts possible.

↑ This chilling 1934 Hollerith poster bears the caption 'See everything with Hollerith punchcards'

COMPUTER PARTITION

While the crimes of the Nazis might stand alone in their brutality, history offers other examples of regimes who used computers for unsavoury purposes. The South African government made extensive use of computers under apartheid to keep track of its citizens and help enforce the division between racial groups.

An arms embargo was enforced upon South Africa from 1977, but computers were still sold for years after. In 1980 a UN Committee told the Security Council that the export of computers should be prohibited. In 1985 and 1986 the Security Council and EU both halted exports of computers for police or military use. However, a total embargo wasn't enforced, so the ruling had little effect on government procurement of computers.

Although it's hard to link computers directly to acts of violence against citizens, computers were certainly used to control their movement and restrict their human rights. Black citizens were each given passbooks detailing where they could go, live and work, and these were tied into a computerised population register for easy reference.

A further possible example is Iraq under Saddam Hussein which, according to an unsubstantiated December 2000 story on the conservative World Net Daily website, once sought to build a supercomputer from 4,000 Sony PlayStation 2 consoles. Quotes attributed in the article to a 'military intelligence officer who declined to be identified' read more like the exaggerations of a PlayStation marketer, however, focusing on 'staggering' graphics capabilities that were "roughly 15 times more powerful than the graphics cards found in most PCs".

Despite the story's dubious feel, it is likely that Saddam's regime would have been more dangerous were it not for the embargo on buying more conventional computing power. It's certainly true that today, consumer PCs with multicore processors and massively parallel graphics processors can be combined in distributed computing projects such as Folding@home to tackle the most complex of problems. The world's most powerful computer, the Cray Titan based in Oak Ridge, Tennessee, owes its supremacy to a 2012 upgrade that, among other things, installed 18,688 Nvidia Tesla K20 GPUs.

EVERYDAY EVILS

While supercomputers (of whatever era) are often used for shadowy purposes, anyone's PC could be co-opted to act maliciously. There are many examples of malware that turn computers into a botnet; a group of distributed computers under the control of a hacker, activist or sometimes even the agents of a state. While botnets don't usually offer much computing power, a large botnet can flood a website or online service with data requests, overwhelming its ability to respond and temporarily preventing the service's legitimate use – a tactic known as a distributed denial-of-service (DDoS) attack.

While many such attacks are certainly criminal or malicious, some highly targeted examples might be considered 'evil'. In the South Korean by-elections of April 2011, for example, DDoS attacks targeted the websites of the National Election Commission and of mayoral candidate Park Won-soon, making it harder for the electorate to look up details of where and when to vote and, potentially, influencing the turnout and outcome of the election. Police later arrested the secretary of the Grand National Party and four others in association with the attacks.

In recent times, the comparative ease with which a DDoS attack can be mounted has helped it to become a tool with which

↑ The Cray Titan, currently the world's most powerful computer, has a massively parallel architecture built from surprisingly mainstream CPU and GPUs

activists can attack the institutions with which they disagree, whether the targets be commercial or political. One example is the ongoing DDoS, hacking and other attacks by pro-Israeli and pro-Palestinian groups that, at the beginning of 2012, resulted in the downing of the Tel Aviv stock exchange, First International Bank of Israel and Israeli national carrier El Al websites, followed by the retaliatory taking down of the Saudi and UAE stock exchange websites.

Such attacks are deliberate, but a website can be overwhelmed by genuine demand, and a service can be swamped as a result of a bug in internet hardware such as a router. Such bugs, or simple mistakes, can quite often be the root cause of computer behaviour that, to the casual observer, might seem malicious. As an example, a simple data entry mistake could result in a black mark on a customer's credit score that subsequently prevents them obtaining another service for which they should in fact be eligible.

Such mistakes are routinely made, but may not be so easy to correct. In November 2012, the financial services company Prudential was fined £50,000 by the Information Commissioner's Office (ICO) in a case where the records of two customers had been mistakenly merged. The mistake, originally made by one of the customer's financial advisors, was understandable as the two customers shared the same forename, surname and date of birth, but the fine arose because Prudential failed to investigate properly when told of the problem.

In every area where modern technology gathers, stores and shares information about us there's the potential for such mistakes, but there's also the potential for deliberate exploitation. In Google's early years – a company whose entire reason for being is to 'organise the world's information' – its staff recognised this threat, adopting the informal motto 'Don't be evil'. It's still referenced prominently in the company's code of conduct, although critics might question the extent to which it influences behaviour.

Companies aren't the only organisations that gather data, with governments across the world eager to retain their grasp on citizens' communications and activities as they use new tools such as social networks. In the worst cases, technology delivers new tools for potential oppression and suppression, from facial or number-plate recognition and tracking in CCTV networks, to censorship or blocking of the web and other services.

WAR MACHINES

There's a limit to the damage that can be done through the gathering and analysis of information or by simple mistakes, but the same isn't true of computer systems that are designed to act on shadowy information or to do harm in the first place. Weapons technology didn't stop with the development of the first ballistic computers; modern warfare relies on a plethora of computerised systems that help map the battlefield, locate and identify friendly troops and enemy targets and, ideally, destroy only the latter. Some, such as GPS, indubitably have far-reaching and peaceful applications, while others such as missile guidance systems may be more specialised.

DRONES APPEAL TO THE MILITARY BECAUSE THEY'RE CHEAPER THAN AN AEROPLANE, AND CAN BE DEPLOYED IN DANGEROUS OR ILLEGAL MISSIONS WITHOUT RISKING A PILOT'S LIFE

We often hear of 'pinpoint', 'surgical' or 'targeted' strikes in the context of military action, but even the most accurate decisions and infallible targeting are only as good as the information on which they're based. When the US declared war on Iraq in March 2003, it launched a cruise missile strike against a supposed leadership bunker and other targets in the hope of wiping out Saddam Hussein and his command, yet the objectives weren't met. Iraqi sources claimed that non-military targets had been hit and civilians wounded, while CBS later reported that the bunker had never existed.

In the past decade or so, the US in particular has intensified its use of unmanned aerial vehicles (UAVs), colloquially referred to as drones, for surveillance and air strikes, both within theatres of war such as Afghanistan, and outside such as in Pakistan. Drones are appealing to security agencies and the military because they're cheaper than an aeroplane, and can be deployed in dangerous or illegal missions without risking a pilot's life, or the difficulties should they be shot down and held captive. However, by reducing human involvement in the gathering of intelligence data and offensive missions that rely on it, many argue that unmanned vehicles increase the risk that innocent people will be killed.

It's often difficult to verify casualty reports from regions in which drones are used offensively, but there are numerous reports of civilians being caught up in supposedly highly targeted strikes. Among the 3,000 people estimated by the Bureau of Investigative Journalism to have lost their lives since 2004 in drone strikes within Pakistan, it's reported that civilian casualties number between 473 and 889. Other estimates are far more pessimistic. Writing for the Washington-based Brookings think tank in July 2009, Middle East security expert Daniel L Byman estimated that for every militant killed by drone strikes, 10 civilians might also lose their lives.

NAUGHTY BY NATURE?

Whatever the exact figures, it's debatable whether weapons of war are inherently evil while they're under the control of humans, who bear the moral and legal responsibility for their use. However, drone technology has improved, and the US Air Force believes that "advances in artificial intelligence (AI)… will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input". In other words, a future generation of drones might decide for itself who to kill.

There's clearly great risk in such a situation. "Military robots are potentially

A shortage of skilled operators could help drive the development of more autonomous weapons

indiscriminate," cautions Patrick Lin, a Stanford University researcher quoted by US news website the Global Post. "They have a difficult time identifying people as well as contexts; for instance, whether a group of people are at a political rally or wedding celebration." Weapons and AI researchers caution that there is no plan for humans to be totally removed from the process, but the military currently doesn't have enough trained operators to meet the demand for UAV sorties, so increased automation would certainly be a great benefit.

While fully autonomous drones might be considered evil – especially if in practice they prove to be less discriminate than human-piloted weapons – in reality even these weapons can't truly be evil without the intelligence, consciousness and morals of a human being. In all of the examples we've looked at so far, where evil has been done it's come from those who designed or used the technology rather than the technology itself, but with computers increasingly able to 'think' for themselves, will this always be the case?

Artificial intelligence is still a distance away from the super-intelligent systems envisaged by computer scientists and writers, but these may still be closer than we'd think. Computer brains may not be able to tackle the reasoning, thought, adaptability and self-learning of the human mind, but for some time they've been able to beat humans at highly specific tasks, such as preventing a car's wheels locking during hard braking or playing chess. More recently the best artificial systems have begun to outperform humans at more complex tasks such as facial recognition, and progress continues.

GHOST IN THE MACHINE

While it's uncertain whether we'll ever succeed in modelling the exact nature of the human brain, it's highly likely that we will manage to create a machine with a similar level of intelligence and, ultimately, a computer that's substantially more intelligent than us. This event is the basis for the concept of 'singularity' in the field of artificial intelligence; a scenario in which mankind creates an intelligent machine that's more capable than we are of designing subsequent intelligent machines. These in turn will create computers that are an order of magnitude more clever, and so on, leading to a sudden and – potentially – unlimited explosion in the intellect and utility of computers.

Such a scenario raises some astonishing possibilities. With unlimited intelligence, future computers could be used to solve problems that have so far defeated humans, such as curing disease, inventing a safe and limitless power source or theorising a new physical model for the universe that incorporates particles, gravity and all the other observed phenomena. They could even tackle vexing philosophical problems such as the existence or otherwise of God, or the meaning of life itself – a scenario anticipated by Douglas Adams in *The Hitchhiker's Guide to the Galaxy*, where the computer Deep Thought designs Earth, the computer, to devise the ultimate question.

A more earthly concern is that, while the tipping point for an AI singularity doesn't require an artificial intelligence similar to our own, it's quite probable that something similar will arise at some point after singularity is reached. This raises the possibility that computers could come to 'think' or be conscious in a similar sense to us, and to understand morality and the concepts of good or evil for the first time. Philosophically, the actions of computers with such an understanding could, finally, truly be said to be good or evil.

It's an intriguing concept, disquieting for some, but even more thorny is the thought that a machine morality borne of a different intelligence and consciousness to ours is likely not only to have different interests, but to have a different concept of morality. In other words, a computer with nothing but good intentions could prove incredibly evil by our standards, because its interests, morality and thus its understanding of evil wouldn't reflect our own.

↑ The General Atomics MQ-1 Predator, the primary UAV used for offensive operations by the US in Afghanistan and Pakistan

SCARE STORIES: WRITERS' FASCINATION WITH THE EVIL COMPUTER

Many developments in technology have provoked suspicion, fear or outright hostility in people whose livelihoods or privacy they've threatened; perhaps most famously when automated textile looms provoked the machine-smashing uprisings of the Luddites in the early 19th century. But there are many examples, too, of those looking further ahead and envisaging more far-reaching changes and threats. Science fiction is liberally peppered with machines, computers, robots and other systems that behave badly towards humanity, from the truculence of HAL 9000 – the ship's computer in Arthur C Clarke's *2001: A Space Odyssey* – to the humanity-ending zeal of Skynet in the *Terminator* series of films.

Such stories are easily dismissed as naive fantasies, written during a more simple age, but while robots and computers have doubtless been the subject of many a sci-fi pot-boiler, many authors have approached computers and their capacity for evil from a serious philosophical viewpoint. HAL 9000, for example, isn't a straightforwardly evil computer, but rather a computer that's struggling to resolve two conflicting instructions: at once he must relay accurate information to the crew members of the Discovery One spacecraft, yet not reveal to them the exact nature of their mission. Unfortunately for the crew, he resolves that the best way to balance the instructions is to fabricate their accidental deaths.

← The evil actions of HAL 9000 arose from a conflict in the instructions given to him by humans

Isaac Asimov was, perhaps, the author best known for exploring such dilemmas, most notably through his robot stories and the three rules of robotics that are attributed to him (in fact, Asimov held that he arrived at them jointly with friend and fellow author Randall Garrett). Many of Asimov's stories explored the conflicts faced by artificial intelligence as it attempted to obey the seemingly simple, immutable and inviolable rules, and the moral and philosophical questions that doing so or failing to do so raised. In the story *Little Lost Robot*, for example, where some robots are created with a truncated first rule that no longer compels them to act to protect humans, Asimov explored the possibility that with such a modification a robot could begin an action that it knew would injure a human, but no longer be compelled to stop it and prevent that harm actually happening.

Such rules might seem a convenient device for fiction, but they're also a plausible solution to what may become a real problem: the need to protect ourselves and the environment from technology that's smarter, faster and stronger than us. The three rules have made their way from stories into serious debate about artificial intelligence, and a similar code may well underscore any future intelligence we create.

ASIMOV'S THREE LAWS OF ROBOTICS

- A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- A robot must obey orders given to it by human beings, except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

ABSOLUTE POWER

Asked if there would ever be a computer as intelligent as humans, US author and singularity proponent Vernor Vinge replied: "Yes. But only briefly." Even in a scenario where technology remains benign, there exists the possibility for mankind to lose control of it and ultimately face competition for energy and materials from, for want of a better word, a species of our own creation.

While such an outcome sounds far-fetched, it's realistic enough that it's now coming to be considered quite seriously. In November 2012, such concerns led to the formation at the University of Cambridge of the Centre for the Study of Existential Risk (CSER), specifically to consider 'extinction-level' risks posed to humans by their own technology.

Writing jointly on the Australian academic website The Conversation, CSER founders Jaan Tallinn and Huw Price likened the prospect of uncontained singularity to a ticking bomb. On containing the risk, they wrote: "A good first step… would be to stop treating intelligent machines as the stuff of science fiction, and start thinking of them as a part of the reality that we or our descendants may actually confront, sooner or later. Once we put such a future on the agenda we can begin some serious research about ways to ensure outsourcing intelligence to machines would be safe and beneficial, from our point of view."

Some academics have suggested that a potential strategy by which we could achieve this is to create only human-based AI, which will share our human values and thus be likely to share and protect our interests. A key problem here is that it seems unlikely that the first super-intelligent AI we create will be similar to ours, and it's by no means certain that we'll ever duplicate the exact nature of the human brain.

An alternative argument espoused by Tallinn and others is to limit artificial intelligences to narrow domains, such that AI can never reach the generalised super-intelligence that would in all likelihood be necessary to displace humans. Certainly, this approach appears more feasible against the background of our current progress, which has delivered super-human ability only in very narrow applications.

TOOLS OF THE TRADE

Computers are probably mankind's greatest tools and, like other tools, the purposes for which we wield them can be good, neutral or evil. They usually do our bidding, and where they don't it's usually by an accident of our design. Either way, the moral responsibility is with us. The future promises increased intelligence and autonomy, however, and the prospect that computers may evolve beyond our control. In such a scenario artificial intelligence may act according to its own morality. If we fail to ensure that this is aligned with ours, we may deliberately or otherwise unleash the first truly evil computers. CS

FURTHER INFORMATION

- The Singularity: A Philosophical Analysis, by David J Chalmers (*http://consc.net/papers/singularity.pdf*)