Computer says no: understanding digital authority

Weggemans, D.J.

2 DECISIONS IN A DIGITOPE

In 1936, Alan Turing invented what is now often considered the first modern computer.[23] Like the famous Turing machine, the computers that followed were typically used for solving complex numerical problems. They were extremely expensive, often broke down and easily filled a large room.[24] And while these colossal machines caught some journalists' attention, their function remained largely unknown to the broader public. At that time, interest and expertise was confined to expert communities within the university and the military. In the early 1940s, Thomas J. Watson, the founder and president of International Business Machines (commonly known as IBM), pessimistically dismissed the idea of commercially produced computers. He reputedly spoke of "a world market for maybe five computers" (c.f. Levy, 2007). How things have changed! Over the last decades computers have conquered the world and permeated our daily lives. In leisure, work and learning they have become commonplace, and many have gained the opportunity to interact with objects that now fit into our pockets but have processing powers that go far beyond that of the supercomputers from not so long ago.

This chapter paints some of the contextual background of this book and portrays the increased role of digital technologies in our world and in human decision-making. The first paragraph (2.1) describes how our living habit is developing in a *digitope* where we are surrounded by digital technologies on which we increasingly depend. In paragraph 2.2, I argue that digital technologies have become embedded in many vital societal structures and organisational operations – also referred to as the emergence of *digital governance.* While this involves many different aspects and activities, our focus is primarily the use of digital technologies to register, store, edit and handle (client) data to foster security (*digital security governance*) and the interactions between technologies and human beings. Here, digital technologies are not merely used to implement decisions but are also, at some point, called upon by professionals when making decisions. This book is not so much about computers, but more about people and how they act and react in response to computer output when making decisions. In paragraph 2.3, I describe the idea of *cognitive enhancement* and the belief that digital technologies are often expected to serve decision-makers by extending their capacities to acquire and process relevant information. It is assumed that with the help of digital technologies we can reach better decisions. While this often proves to be the case, the first thought for writing this book was triggered by those stories where people are seriously led astray by digital technologies. Paragraph 2.4, in turn, further grapples with this potential drawback and discusses the common anomalies associated with digital technologies' involvement in human decision-making predicaments. In short, I explain why it is not always a good idea to blindly rely on these technologies.

2.1 A WORLD OF DIGITAL TECHNOLOGIES

We live in a highly technological world. Philosophers claim that the biotope in which we used to live has transformed into a *technotope* (De Mul & Van den Berg, 2011; De Mul, 2013; Swierstra, 2013; Van Boeckel, 2015; Mendes, 2021). Jacques Ellul (1962) described it as "the new and specific milieu in which man is required to exist, one which has supplemented the old milieu, viz., that of nature" (Ellul, 1962: 394). As described elsewhere, technology has become our environment - the *technotope* is not just the place where we live, but also makes living possible and shapes our existence (Van Boeckel, 2015: 215; Ellul, 1964: 325; Ellul, 1990: 15). We have created technological environments beyond which we cannot survive (De Mul, 2013: 39; Kockelkoren, 2007: 4). But perhaps we should refine this analogy and describe our world as more and more resembling a *digitope:* an environment that is pervasively saturated with digital technologies. For many things vital to our existence, we increasingly depend on digital technologies – whether it is for food production, traffic coordination, water management, financial systems, and security. It seems that that modern people not only lack the choice to not live in a technological environment, but that they can also hardly escape the web of digital technologies that surrounds them. In this sense, one can say that digital technologies increasingly facilitate modern life.

In his book *Future Politics* (2018), Jamie Susskind also describes this world of digital technologies. He writes that in a matter of decades, digital technologies have become more *capable*, *connected,* and have generated a world of *information abundance*. These intertwined developments have been important for the transformation of our world and the many decision-making contexts we find ourselves in. Before further addressing the role of digital technologies in todays and tomorrows (security) organisations, I first briefly discuss these underlying developments. I do not presume this discussion to be complete. There is simply too much going on. Rather, the purpose here is to gain a general understanding of the developments that allow these digital technologies to become part and parcel in our world. Let us start with the observation that digital technologies are now capable of things that were unimaginable not so long ago.

2.1.1 Capable Technologies

In April 1965, Gordon Moore published an essay in the weekly journal *Electronics*. Moore, who later co-founded high-tech company Intel, envisaged the still-new technology of integrated electronics leading to wonders such as electronic wristwatches, smart vehicles, home computers and "personal portable communications equipment" (1965: 82). But it was one particular prediction about the future development of microelectronics this essay became famous for: Moore predicted that the number of transistors on a chip would double every year for at least a decade, which generally meant that computer processing power would too. Moore's law (as it became known) turned out to hold for over half a century. Moreover, he expected that the dramatic growth in computing power would be accompanied by an equally dramatic decrease in costs. In the late 1950s, a single transistor of the size of a cotton fiber cost about $10 (Kilby, 2001: 487). Today, they cost less than a cent and some transistors have become so tiny that they are invisible to the naked eye. Millions of transistors can now be packed onto a chip the size of fingernail (IBM, 2017). This has made digital technologies accessible to nearly everyone.

Meanwhile, some have argued that Moore's law is nearing its endpoint. Technologists say it is taking increased effort to get transistor sizes down, and costs are bound to get higher (Waldrop, 2016). Eventually it will just become physically impossible to squeeze any more transistors into the same small area – things only get so minuscule before they start breaking down. At the same time, innovative technologies such as 3D chips and quantum computing are currently being developed with the promise of even greater computing capabilities (DeBenedictis, et al., 2017). The development of computers is one of both evolution and revolution. As Susskind puts it, "[it] is a story of a succession of increasingly powerful methods of processing information, each developing exponentially, reaching its physical limitations, and then being replaced by something better" (2018: 40).

It suffices to conclude here that over the course of the last several decades computing power – the rate at which computers can execute certain standard tasks, as Nordhaus defines it (2007: 134) – has improved tremendously. This has led to machines that are overwhelmingly better than anything before. Today's smartwatches are at least twice as fast as the 1985 supercomputer Cray-2 (Exchange, 2016). The cutting-edge computer that landed Apollo 11 on the moon is about as powerful as a contemporary pocket calculator. And computing power is not the only thing developing like this. Think, for example, of network capacity, memory capacity and sensing technologies that have also evolved spectacularly.

With computing products improving in terms of processing power, costs and size, new digital applications and technologies have been on the rise. One example is the field of Artificial Intelligence (AI).[25] In broad terms, the term refers to non-humans (generally assumed to be computers and robots) displaying behavior that is associated with intelligent beings. Matthew Scherer defines AI as "machines that are capable of performing tasks that, if performed by a human, would be said to require intelligence" (2016: 362). AI is an exciting field and holds many promises. Google CEO Sundar Pichai (2018) called AI "probably the most important thing humanity has ever worked on (..) more profound than electricity or fire". Much of the interest, however, has focused on the development of *generalized AI*.[26] That is, the idea that there could be systems or devices with the full range of human intelligence and the ability to perform any human task (Kumar & Thakur, 2012: 57; Kurzweil, 2005). These machines then will be able to reason, think, learn, and do the same things as human beings do. While such generalised AI has infiltrated mainstream culture via Hollywood and is a focus of a lively philosophical debate, the world is still nowhere near creating this form of AI and the question remains if it ever will be. In contrast, *applied AI* is everywhere around us.[27] These are software applications capable of specific, often quite narrow, problem solving or reasoning tasks which have become deeply embedded in a wide range of fields and activities. Autonomous vehicles use AI to recognise people and objects and take the right decision on different occasions. In healthcare, AI systems are used for medically diagnosing and predicting diseases. Chatbots write student papers, provide costumer services, as well as therapy to countless people around the world (JMIR, 2024). Businesses use AI for recommending a particular product. We can easily overlook how commonplace this form of AI has become, as is also summed up nicely by Swedish philosopher Nick Bostrom (2006) who stated in an interview that: "a lot of cutting edge AI has filtered into general applications, often without being called AI because once something becomes useful enough and common enough it's not labelled AI anymore".

So, while on the one hand the vast advances of computing have not captured human intelligence over wider domains (and maybe never will), computer programs have been developed, on the other hand, that are capable to operate autonomously and perform specific tasks just like humans would. These and other developments, such as in the field of robotics or sensing, have created a world filled with increasingly sophisticated machines.

2.1.2 Connected Technologies

A second important development is that the rise of the internet and computer networks. It is difficult to establish when the internet exactly began – among other things because there is still considerable disagreement over what the internet is in the first place. However, one date that is frequently mentioned is 29 October 1969. On that day, as internet pioneer Leonard Kleinrock writes, "the internet came to life when it uttered its first words" (Kleinrock, 2004: 195). These first words, sent over a wire between two computers several hundred miles apart, were not intended to be as considerate as that of Samuel Morse when he wrote the first telegraph message in 1844: "What hath God wrought?" (Rainie & Wellman, 2012: 59). Instead, in a crowded room, engineering student Charley Kline was tasked to login from a laboratory at UCLA to a Stanford machine. He managed to transmit the "L" and the "O", but when Kline typed a "G" the system crashed, and the connection was lost. Nevertheless, the first message on the nascent internet turned out to be, although by accident, fittingly profound: “LO” (Kleinrock, 2004: 195).[28]

At that time, future expectations of the internet (then known as the ARPANET) were already high. In a press release that introduced the internet to the public, Kleinrock articulated his vision of what the internet would become:

> “As of now, computer networks are still in their infancy. But as they grow up and become more sophisticated, we will probably see the spread of ‘computer utilities’ which, like present electric and telephone utilities, will service individual homes and offices across the country.” (Kleinrock, 2004: 195:6)

With the advent of these digital networks, the sharing of files and messages began. In the early 1970s, engineer Ray Tomlinson, who worked for a U.S. Defense Department contractor, came up with a way to send messages to users on other computers. This resulted in the first email that was sent from one computer to another situated right beside it. Again, the message was not intended to be particularly philosophical. In an interview with *The Wall Street Journal* (2016), Tomlinson said the message was “entirely forgettable” – something like “QWERTYUIOP”.

In its early years the internet was largely confined to universities. But with the rise of the personal computer in the 1980s and early 1990s, many individuals and commercial organisations also got connected to the internet. As Rainie and Wellman conclude: “where the 1980s were the decade of stand-alone computers, the 1990s became the decade of connected computers” (Rainie & Welllman, 2012: 61). They explain that after the first easy-to-use browser became available, the web attracted an audience of 50 million people within 4 years – it took radio thirty-eight and television thirteen years to amass that many (ibid).

In the years that followed, the internet developed in a similar way as computing power and processor prices. Broadband technology replaced the slow dial-up access where a connection to the internet was established by dialing a specific number on a conventional telephone line. With the introduction of broadband connections, the internet would always be “on” and much faster. As a result, users’ internet connection speed (or bandwidth) roughly doubled every 21 months. As Nielsen (2023) has shown, bandwidth has increased from 300 bits per second in 1984 to more than 1 gigabyte per second in 2020. At the same time, moving all these bits and bytes has become 50 percent cheaper every year (Rainie & Wellman, 2012: 63).

The internet became mainstream in the 21st century. In 2018 the International Telecommunication Union (a specialised agency of the United Nations that is responsible for issues that concern ICT) estimated that more than half of the global population, or 3.9 billion people, is now online (ITU, 2018: 2). In developed countries, four out of five people use the internet.[29] And even when someone may want to avoid using the internet, it is increasingly difficult to do so.[30] The internet is getting increasingly *pervasive*, Susskind concludes (2018: 43). More and more things are now connected and so many daily activities are linked to the internet. In today's world people are often online even when they never sit down at a computer or hold a smartphone. Many objects, ranging from toothbrushes to door locks and from thermostats to parking spaces have become embedded with processors, sensors, software, and internet connections. Cisco Systems estimated that now there are over 50 billion interconnected devices in use and that eventually 99 percent of the physical objects will be connected to a network (Greengard, 2015: 13). This ever-expending web of internet-connected devices is referred to as the "Internet of Things" (IoT) (Gubbi, Buyya, Marusic & Palaniswami, 2013; Greengard, 2015).[31] Not only anyone, but almost anything can now be online.

In short, the rise and growth of the internet and computer networks has made the world more connected than ever before and this allows for the continuous exchange of information between people, between machines and between people and machines (Susskind, 2018: 44). This has typically led to the availability of large-scale datasets, to which we will turn now.

2.1.3 Information Everywhere

Our digitope contains a tremendous and ever-growing amount of digital information.[32] More and more of our actions, interests, relationships, movements and communications are being turned into data. Similarly, data is increasingly amassed on the operations of machines as well as on our built and natural surroundings. Cukier and Mayer-Schonberger speak of the datafication of the world (2013:78). And all this data can be used for a variety of purposes, one of them being the making of decisions.

It is difficult to grasp the amount of information in the world and its growth. Martin Hilbert and Priscilla López (2011; 2012a, 2012b) tried to quantify the world's entire capacity to store, communicate and compute information. They conducted an extensive study and gathered both analogue and digital sources, including books and newspapers, photographs, X-rays and paintings, hard disks, vinyl records, floppy disks, CDs and DVDs, videogame consoles, pocket calculators, mobile phones and car navigation systems, phone-calls and internet subscriptions, radio, and television broadcasts and even credit card chips (Hilbert, 2012; Hilbert & López, 2011, 2012a; 2012b). They estimated that the amount of stored information grew from 2.6 exabytes in 1986 to more than 300 exabytes by 2007 (2012a: 958). To put this astronomical amount in more practical terms: a two-hour movie or roughly 50,000 typed, double spaced pages of text are about one gigabyte. And one exabyte equals one billion gigabytes. In short: a lot of information (Schonberger & Cukier: 8).

Today most of our information is digital. While in 1986, only 99 percent of our data was analogue (books, papers, printed photographs, cassette tapes, and the like), this quickly changed and by the year 2000 already a quarter of the stored information in the world was in digital format (Hilbert, 2012: 9). However, by 2007, 97 percent of all data was already in digital form (ibid; Kitchin, 2014: 85).

While analogue information barely increases, the amount of digital information in the world further explodes. As was the case with computing power and internet connections, the speed with which we produce information is said to be doubling every two years (Kitchin, 2014: 69).[33] We now generate the same amount of information in a few hours as all humans have done from the beginning of civilisation until 2003 (Susskind & Susskind, 2015: 161). It is expected that in the coming years the amount of digital data that is generated annually will be around 44 zettabytes (IDC, 2014) - 44 trillion gigabytes or 44,000 exabytes.[34] As one study characterises it: if all this data was placed on tablets and stacked up, they would have stretched two-thirds to the moon in 2003 (when the amount of stored information in the world was estimated around 4,4 zettabytes) and 44 zettabytes would equal more than 6,5 complete stacks from the earth to the moon (ibid:2). In 1949 Claude Shannon, the inventor of the bit, estimated that the US Library of Congress - the largest collection of information he could think of – was about 12500 MB (Hilbert, 2012: 8). Our digital world is the equivalent of providing every person alive today 500 times the information that was stored in Shannon's US Library of Congress.[35]

There are various reasons for this information explosion. The most obvious one, Kenneth Cukier (2010) writes, is the emergence of more capable and cheaper technologies with sensors and gadgets digitising lots of information that was previously unavailable (2010: 2).[36] The fact that digital technologies have become ubiquitous, and many activities are now undertaken by or near them, has resulted in large amounts of data that can be gathered without much effort.

Not only is there more data to collect, but much more data can also be stored against lower costs. Within decades, organisations moved from using punched cards to tape and later to hard disks and the cloud for storage (Cortada, 2008:146). Over the past half- century, the cost of digital storage has halved roughly every two years while increased in density "50-million-fold" (Mayer-Schonberger & Cukier, 2014: 101). Kitchin concludes that: "costs have now become so cheap that it is possible to store enormous amounts of data for negligible costs, with the deletion of old files virtually unnecessary" (2014: 85). By the same token, it has become possible to copy and reproduce digital information many times very cheaply (Susskind, 2018: 62).

Last, all the computational power described in the previous section allows us to actually process all the data out there. It has allowed data scientists to develop techniques to clean, organise and analyse all this data (Broeders et al., 2016:45). The promise that actionable knowledge can be discovered makes that all data now is regarded as valuable and something worth collecting (Mayer-Schonberger & Cukier, 2014: 100).

In sum, the previous sections illustrate that, over the last decades, computing, networks, and the availability of information have all developed at a staggering pace. These developments underpin the world of digital technologies – a world that is soaked with information and capable, connected machines. In this world, individuals, governments, and businesses all have come to rely on digital technologies for effective decision-making. Let us now turn to this and the way these technologies facilitate decisions on security and beyond.

2.2 DIGITAL GOVERNANCE AND DECISION-MAKING

The digitisation of our world has not only revolutionised our personal but also organisational life (Milakovich, 2012; Dunleavy et al., 2006). Many branches of government and businesses have embraced digital technologies to reshape how theyfunction and are organised (Kitchin, 2014: 117). The term *digital governance* broadly refers to the use of digital technologies in the furtherance of organisational goals.[37] Digital governance is an umbrella term, encompassing a wide variety of phenomena. For example, it is used to describe the importance of digital systems in facilitating greater coordination and control *within* organisations as well as their role in delivering services and interactions *to* clients or society at large. In this latter context, digital governance was initially also a synonym for e-governance to highlight the utilisation of the Web for delivering information and communicating with individuals.

Meanwhile, the term is mostly associated with the use of digital technologies in organisations for information handling and decision-making.[38] For example, amending Webers ideas about bureaucracy, Zuurmond (1998) speaks of the emergence of the *infocracy* – public organisations that revolve around digital technologies for the collection of data, sharing of information, and (semi-)automated decision-making. Digital technologies have become part and parcel to gather, process, store, edit and distribute information in almost all public and private organisations over the past 60 years (Kitchin, 2014: 117). Common examples are (variations of) databases, Decision Support Systems (DSS), expert systems and automated decision-making systems (Busch & Henriksen, 2018: 3-4). These are diverse in both the techniques they deploy as well as in their applications, and it goes beyond the scope of this book to detail them further. What is of interest here is that all these digital technologies in some way interact with human decision-making processes.

One could place these digital decision-making technologies on a continuum (see figure 1 below).[39] At one end of this continuum are those systems that only provide information on particular affairs – think of databases that inform police officers on the criminal history of an individual or a medical expert system that provides a diagnostician with potential diseases that might underlie a patient's symptoms. It is up to the decision-maker to interpret this information and to put it into practice. In the middle of the continuum, we find digital systems that also provide decision-makers with courses of action. One example is navigation systems: they interpret the world for drivers and provide them with clear instructions on where to go next. Similarly, there are algorithms that advise bank clerks in credit decisions, that tell managers who to hire or that produce risk-scores and advise security professionals on how to act on this (e.g., invite for further questioning or deny entry). At the other end of the continuum are those information systems that take over (initial) decision-making process in their entirety. For example, most traffic violations are independently handled by computers. Similarly, many government decisions about social benefits and taxes are often made by computers (Van Eck, 2018; Meijer et al, 2021: 838). It is only when someone appeals this decision that a human-decision makers gets involved.

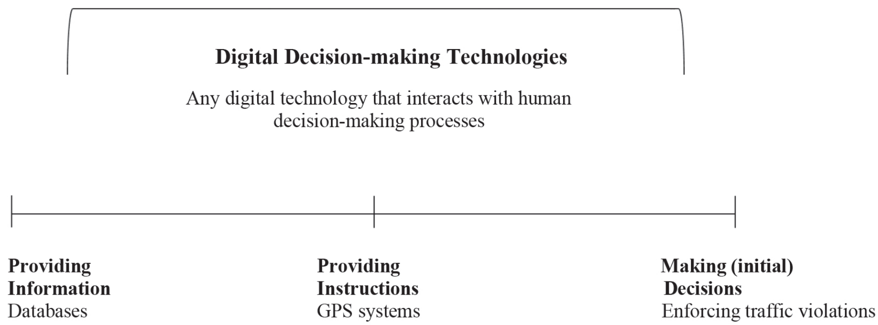

Figure. 1 Digital technologies and human-decision making

In their seminal article Bovens and Zouridis (2002) describe how information systems can transform the structure of organisations.[40] They argue that face-to-face contacts are gradually diminishing, and decision-making procedures are more and more routinised by computers. In this vision, human decision-makers are increasingly connected to the organisation via the computer, electronic forms with fixed templates are used to fill in client data and knowledge management systems and digital decision trees have steadily reduced the scope of professionals' discretion.[41] Ultimately, human-decision makers could be replaced entirely by fully automated systems that make decisions about clients based on collected data and pre-programmed algorithms (ibid: 180). According to Bovens and Zouridis (2002: 175), because of the rise of digital technologies it is more and more appropriate to think of many executive organisations as "rooms filled with softly humming services, dotted here and there with a system manager behind a screen" (2002:175).

For the purpose of this book, two points are relevant here. First, as Bovens and Zouridis themselves acknowledge, we must be careful to generalise this latter mode of transformation. These authors are describing large "decision-making factories" that handle thousands of formalised transactions. Here the automation of decisions is particularly likely because of their if/then structure, which can easily be translated into algorithms and decision-trees. However, in many organisations decisions are not as structured and hence it is unlikely that similar transformations that leave little or no room for human decision-makers can be observed in non-legal, non-routine interactions with clients (e.g., teaching, nursing and policing) (Bovens & Zouridis, 2002: 180). There are still many services that cannot be provided by algorithms alone (Buffat, 2015: 153).

Second, even when decisions can be mapped out entirely in digital systems, humans will always be "in the loop" somewhere. For example, public tax agencies use software to automatically process tax reports. But reports that are flagged by the algorithm are, in turn, subjected to human scrutiny. Similarly, with automated systems for traffic enforcement, a specialised official will get involved when someone submits a notice of objection. When a particular case emits some kind of signal, there are always human officials that will enter the picture and make a decision. Ergo, even when information systems are at the one extreme of the continuum, in which they start making decisions themselves, this does not mean there is nothing left for human decision-makers to do. Rather, their place in the decision-making process generally shifts from making the initial decision to evaluating it.

What I am interested in this book is not so much the extent to which digital technologies supplementan decision-making activities or even primarily how digital technologies are transforming the general structure of organisations. Rather, I take as a starting point the observation that in today's (and tomorrow's) organisations many decisions result from the interplay that has emerged between humans and digital technologies – technologies that capture, manipulate, analyse, or utilise information, and that communicate insights which are of importance to the user's decision task. In this context I discuss those systems that facilitate human-technology interaction and decision-making. These are generally not large, strategic choices, such as where to locate a new detention facility, but most of the time concern smaller-sized choices that are central to individual's life and work – the everyday decisions, for example on who to investigate, who to grant access to, how to respond to a given situation, etc.

Although digitisation has transformed almost every aspect of our personal and professional world, I take the domain of security as the main point of focus because of its potential impact on human beings, groups and societies at large. Digital technologies have become indispensable for security, law enforcement and the fight against fraud. In the introductory chapter I already briefly introduced the idea of *digital security governance* which refers to the use of digital technologies for the collection, processing, storage, sharing and/or application of digital information for security (Eijkman, 2014: 116). It is not easy to analyse how information systems are used exactly in the field of security. One reason for this is the secrecy that often surrounds security policy. Another reason is that some of the implementations of information systems are still experimental, which comes with reluctance of further public discussion (see also Broeders, 2017: 311). But like in any other part of the modern world, digital systems have obviously become central in security practices world-wide. Security practices are increasingly characterised by risk profiles, large databases, datamining, smart and sensing technologies, etc. And the ongoing technological innovation already hints at a future in which digital technologies are the primary means for security.

The result is that more and more security professionals, working on the streets or from behind desks now operate in close proximity to machines that provide them with information or advice on how to act in a particular situation. Police officers can digitally "run a number plate" to check if a vehicle is stolen, customs agents are alerted by their computer when a passenger is "deemed high-risk", judges are advised by software in making parole decisions, prison officials use computer programs to determine inmates' release date, expert systems help child protection services to make appropriate choices on youth well-being, tax inspectors work with automatically generated red flags for fraud identification, and so forth. Obviously, the digital technologies mentioned here are all very different - for example in where they can be placed on the previously mentioned continuum of digital technological interactions with human decision-making processes. But what they share, however, is that they are there to help professionals to make sensible and effective choices. These technologies communicate useful information about the world so that better decisions can be made. But it is, *ultima ratio*, up to the user to decide whether to act on such a digital depiction of reality. The focus here thus lies on decision-situations in which information systems are involved, but where there is, at least in theory, a user that is ultimately in control. As stated in the introduction: it is easy to concentrate on the technology components (hardware, software, and data) and forget that to fully understand how digital systems work, we must not overlook the people that use them during the course of making a decision. Let us therefore now further focus on those decision-makers and how information systems might boost the quality of their functioning.

2.3 TECHNIQUES TO IMPROVE THE HUMAN MIND

Humans have a finite mind. We cannot know everything or deal with all the information that is coming to us. And sometimes, we fail to recall what we once knew. Or when we do manage to retrieve previously stored information, it is not necessarily completely accurate. Although the human mind can do wonderful things, we are surely limited in our ability to process and store information, as Marakas concludes (1998: 59).

Perhaps the best-known cognitive limitation is what George Miller (1956) referred to as "the magical number seven": we are only capable of keeping about seven chunks of information in our working memory (often called short-term memory). In other words, when making a decision we can only handle a few active items of knowledge. We are also clearly limited by the speed with which our brain can process input from our environment. In his book *Flow,* Mihaly Csikszentmihalyi (1990) estimated that our maximum processing capacity is only about 126 bits per second (c.f. Heylighen, 2002: 12).

These cognitive limitations often prevent us from making the best decisions. To use Herbert Simon's words, people have "bounded rationality" (1997:88-9). Bounded rationality recognises that our cognitive capabilities are often insufficient to find optimal solutions to real-life problems (Mallard, 2016: 2). While we may wish to behave in a perfectly (or objectively) rational way, we are typically unable to achieve this because of the enormous effort involved in such a process (Mussweiler et al, 2006: 34).[42] We cannot go over all the relevant alternatives to a particular problem, and even if we could, we still cannot assimilate all this information for making an appropriate decision (Marakas, 1998: 68). Instead, Simon (1997: 118-120) argued that, bounded by our cognition, we tend to simplify reality and focus on finding solutions that are "good enough" rather than perfect. When a solution meets our preconceived threshold of acceptability, we adopt it and instantly stop looking for a better one. Simon calls this strategy *satisficing*.[43]

For a long time, humans have been searching for ways to overcome the limitations of their minds and enhance their cognitive abilities. Sandberg & Bostrom (2006a; 2006b; 2009) describe cognitive enhancement as the amplification or extension of the core faculties of the mind. These involve all processes involved in the organisation of information, such as *perception* (recognising information), *attention* (selecting relevant information), *remembering* (storing information) and *deciding* (using the information to guide our actions). There are various conventional ideas about how to improve these faculties. Consider activities such as education and training for cognitive development, the consumption of coffee and dietary supplements, practicing martial arts, and engaging in yoga. While some of these methods may be more effective than others, they are all directed at improving our cognitive functioning (Vedder & Klaming, 2010). There are also more controversial methods that seek to achieve this goal, for example so-called "smart-drugs", genetic modification[44] or advanced neurotechnologies such as Transcranial Magnetic Stimulation (TMS)[45]. While the list of such *internal*[46], biological enhancements is steadily expanding, the most dramatic advancements in information processing occur *external* to the user, driven by the development of digital technology (Bostrom & Sandberg, 2009: 311-2).

Humans have always used external objects to work around their cognitive limitations. Think for example of externalising ideas in books, time in calendars, rules in road signs and math in calculators (Sandberg & Bostrom, 2006b). Common artifacts such as pen and paper help us to extend our working memory by storing information on paper or allow us to solve mathematical problems that would otherwise frustrate our unaided brains (Clark, 1999: 349). However, the cognitive function of information systems goes much further. They allow us to do many new things. As Bostrom and Sandberg write, "external hardware and software support now routinely gives human beings effective cognitive abilities that in many respects far outstrip those of biological brains" (Bostrom & Sandberg, 2009: 312). Information systems and other digital applications extend our ability to process information and knowledge - something much welcomed in a society where people are subjected to *information overload* (Heylighen, 2002) - as well as decrease the time associated with reaching a decision (Marakas,1998: 5). Moreover, these technologies can also generate new - previously unavailable - evidence in support of a decision and allow us to handle complex problems we could not even dream of before.

Digital technologies allow for the large-scale pooling of data and the delivering of relevant information to a decision-maker. Depending on the type of digital technology, this output can take the form of a query response, advice, an update on states of affairs, etc. (see figure 1). These technologies support human cognitive processes by storing and providing relevant information to help the human user correctly assess a given situation or problem, and to respond appropriately (Parasuraman & Manzey, 2010: 391). The assumption is that if decision-makers are provided with digital technologies that expand their processing abilities, they are better equipped to cope with complexity and manage uncertainty in a rational way. They could analyse problems more in-depth and as a result make better decisions. In short, following Bostrom and Sandberg (2009) here, a cognitively enhanced person is thus not necessarily somebody with super-human cognitive capacities. Rather, it is somebody who is thought to have benefited from an intervention “that improves the performance of some cognitive subsystem” (Bostrom and Sandberg, 2009:312).

As such, cognitive enhancement is expected to improve the quality of the decisions made, it is therefore no surprise that organisations have very much welcomed these technologies. If it is the limits of our cognition that prevent us from making better decisions, then enhancing decision-making makes great sense. In this view, digital technologies are simple means or tools to achieve our goals. However, this book started with the observation that our cognitive functioning is not necessarily enhanced with digital technology. In the following chapters I describe how decision-makers can develop a strong preference for the communications of a computational device. Instead of being a complementary source of information, digital output may become much more than that. As was also mentioned in the introductory chapter, there is a wealth of empirical research showing that the communications of a computer tend to draw users’ attention and are easily accepted (Parasuraman & Manzey, 2010: 391). The output of an ICT application can come to function as “a heuristic replacement for vigilant information seeking and processing” (Mosier and Skitka, 2009: 205). Mosier and Skitka (1999) refer to this phenomenon as *automation bias*. This idea fits with the observation that humans are cognitive misers that in decision-making tend seek the road of least cognitive effort (Fiske & Taylor, 1994). So instead of including the information that is digitally offered to us in a comprehensive information analysis and evaluation, we use it as a simple heuristic (Tversky & Kahneman, 1973): the digital output functions as a replacement for other (more effortful) processes of information processing (Mosier & Skitka, 2009). For us to understand this, I argue in the following chapters, it is important to start viewing digital systems in relational terms and as an object of interaction rather than a simple tool for enhancing our cognitive functioning. Only then we can further investigate *why* their output can become a heuristic in the first place – a question I answer in this book by arguing digital technologies can become digital authorities. For now, let us first establish that it is not always a good idea to blindly follow the information communicated by a digital technology and that the promise of cognitive enhancement cannot always be met. After all, would it be a problem if a technology vividly shapes what we think and do if it would always be right?

2.4 WHY YOU SHOULD NOT AUTOMATICALLY FOLLOW A MACHINE

Digital technologies tend to come with high hopes and expectations about improving the quality of our decisions. Unfortunately, these expectations are not always met and blindly following these technologies can have adverse or even disastrous effects. While often we recognise in time that the computer's output is sub-optimal or downright wrong and decide not to follow it (see chapter 7), we can also find out when it is already too late. In this book's introduction we already touched upon the obvious examples from the world of navigation systems. Judging by the many news articles, motorists are being led astray by their navigation system all the time. Rather than looking at road signs and the view through the windscreen, people can clearly believe that "the machine knows where it is going!", as Michael Scott yelled in an episode of *The Office,* while driving his car into a lake. The aim of this paragraph is to further understand why information systems might fail to provide us with the expected (correct) output, and thus why blindly following these technologies may turn out to be problematic.

2.4.1 Technology-based anomalies

When using digital technologies, people generally expect them to do the right thing and to do them right. I use the term *anomaly* for those situations in which these technologies fail to achieve this expected outcome.[47] Anomalies may result from a variety of problems in the digital technology itself. Such *technology-based anomalies* occur when there are flaws or inadequacies in the technology itself. Three factors appear to bring about such anomalies: (a) inaccurate data; (b) bad or ill-conceived algorithms; and (c) flawed or deficient hardware. I briefly discuss each in turn.

First, the use of *inaccurate data* generates a situation that does not conform to an expected outcome. In computer science, there is the popular phrase "Garbage in, Garbage out", which refers to the idea that when flawed or nonsense input data is used, it will generate incorrect output. In other words, if someone puts the wrong figures into digital technologies, it will produce the wrong answers. Many data inaccuracies result from wrong initial data entries (see also 2.4.2) – (deliberately or mistakenly) false data is used in the first place.[48] For example, someone is supposed to enter *Sara* but enters *Sarah* instead, puts a correct value in the wrong field, uses information that is made up, or selects the wrong option from a drop-down menu (Olson, 2003: 44). And even when the data was entered correctly, at some point it may no longer be up to date. The use of inaccurate data can have profound consequences, as is painfully highlighted by the case of the Russian activist Aleksandr Dolmatov who fled to the Netherlands in 2012 to avoid arrest. While awaiting political asylum, Dolmatov was unjustly placed in a Dutch detention center following a mistaken entry in the computer system (INDiGO), classifying Dolmatov (and some 300 others) ready for deportation. He committed suicide in his cell.

Second, technology-based anomalies can also occur because of *bad or ill-conceived algorithms*. In essence, an algorithm is nothing more than a set of instructions for solving a problem or accomplishing a task. When these instructions turn out to be wrong - for example because they are based on incorrect assumptions, false logic, skewed training data or programmers that have simply wrongly coded them (O'Neil, 2016; Wachter-Boettcher, 2018) - they will do or tell us the wrong things. For instance, in 2013, a t-shirt company called Solid Gold Bomb was heavily criticised for selling a shirt on Amazon with the phrase "Keep Kalm and Rape a Lot". It turned out that the company had generated the designs by using an algorithm that was randomly matching verbs with adjectives from the English dictionary. Apparently, the possibility of the algorithm generating offensive slogans was not considered and nobody had bothered to check the t-shirts that were offered for sale (example from Dormehl, 2015). However, things get more serious when an ill-conceived algorithm systematically confuses people (including hundreds of toddlers) for potential terrorists, leading to hours of delays at airports and airplane boarding denials – an example that will be further introduced in chapter 4.

Last, technology-based anomalies can be a consequence of some *hardware deficiency*. While digital systems greatly improved over the last decades (see paragraph 2.1), hardware - the physical stuff - did not always manage to keep up with our demands. A case in point are facial recognition systems. More and more public and private organisations world-wide are using this technology for identifying people from digital images or videos. Yet, there are still many stories of smart cameras incorrectly identifying people or failing to recognise them altogether. One of the reasons for this is the poor image quality of the cameras used (Introna & Nissenbaum: 2010).[49] A case in point is a financial advisor who was violently arrested and locked behind bars after facial recognition software recognised him as a bank robber. The charges were dropped later after a human operator discovered that a mole and the length of the man did not match that of the suspect. But the damage was already done. He ended up homeless, losing his job and access to his children (Fry, 2018: 123).[50]

2.4.2 User-based anomalies

An anomaly in the context of this book is not necessarily a problem that arises from the digital technology itself. Anomalies can also be a direct result of human users doing the wrong thing in their interaction with the technology. With *user-based anomalies* it is technically not machines, but human beings who lead themselves astray. Yet, in these situations people are typically unaware of their own role; they are focused on the technology and its output. People experience they are getting information or advice from a computer and when something goes wrong, they can be tempted to - at least in first instance - attribute it to the machine: “the computer said so”.

Psychologist James Reason (1990) has written extensively on the issue of human error. Such errors can be broadly divided into two categories. First, there are *execution failures*[51]: a person intends to act in a certain way, and while the action is appropriate, it is carried out incorrectly, and the desired result is not achieved. Slips and lapses are execution failures. Reason explains that they “result from some failure in the execution and/or storage stage of an action sequence” (Reason, 1990: 9). When the appropriate action is not executed as planned, the error is called a *slip*. The term *lapse* is generally reserved for more covert error forms, involving failures of memory, such as when an action is simply omitted or not carried out (ibid.). Reason argues that many slips and lapses result directly from attention failures (ibid: 69). For example, when using a digital technology, users can inattentively give the wrong input to a digital device and, in turn, will be provided the wrong output. People can fail to notice that they misspell their destination in a navigation system and end up miles from where they wanted to go. A more serious case is that of a 16-year-old patient in a respected San Francisco hospital. Because of a mistaken entry, the computer prescribed the patient not one Septra pill (a common antibiotic), but 38½ of them (Wachter, 2015: 127-130). Even though the nurse judged that this dose was exceptionally high, she did administer the pills to the patient who later suffered a grand mal seizure and stopped breathing. He only just survived the overdose.

Second, there are *planning failures*. Here, a person did what he or she intended to do, but it did not work because the plan or goal was wrong. This type of error is also referred to as a mistake which can be defined as “deficiencies or failures in the judgmental and/or inferential processes involved in the selection of an objective or in the specification of means to achieve it” (Reason, 1990: 9). Mistakes tend to be grounded in unfamiliar situations or a lack of expertise. That is, individuals either inappropriately apply some preestablished plan or problem solution to a new situation, or they do not have an appropriate “off-the-shelf” routine and are forced “to work out a plan of action from first principles, relying on whatever relevant knowledge [they] currently possess”, Reason explains (Reason, 1990: 12-3). Take for example the situation in which someone is working with a new technology and is relying on previous user routines or selectively processes relevant task information. When these routines are wrong or relevant information is not given attention, the odds are that the digital output will also be incorrect.

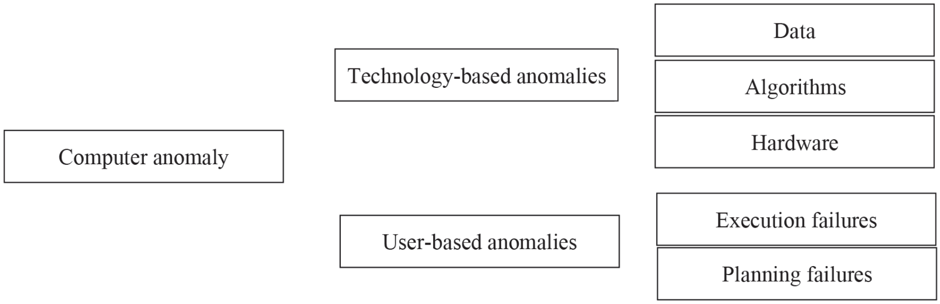

Figure 2. Understanding information system anomalies

The distinction between technology-based and user-based anomalies is artificial in nature and the different categories can (and often do) overlap. What is of importance here is the observation that it can be a bad idea to blindly rely on an information system's output. For different reasons these technologies can fail to meet our expectations and may lead us astray. In these situations, the central issue becomes human critical thinking. But while we may say we are aware of this, there are still many situations in which things go (really) wrong. Even when the information of a computer can easily be determined to be false, we can still decide to act upon it. In the next chapters we look further into this tendency to do what the computer says and to rely on it beyond our own judgement.